5-21-2019

Mapping Relict Charcoal Hearths in the Northeast US Using Deep Learning Convolutional Neural Networks and LIDAR Data

Eli Anderson
eli.egan-anderson@uconn.edu

Mapping Relict Charcoal Hearths in the Northeast US Using Deep Learning Convolutional Neural Networks and LIDAR Data

Eli Anderson

B.S, Cornell University, 2015

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

At the

University of Connecticut

2019

APPROVAL PAGE

Master of Arts Thesis

Mapping Relict Charcoal Hearths in the Northeast US Using Deep Learning

Convolutional Neural Networks and LIDAR Data

Presented by

Eli Anderson, B.S

Advisor_____

William B. Ouimet

Associate _____

Chandi Witharana

Associate _____

Andrew Jolly-Ballantine

University of Connecticut 2019

Acknowledgements

I would like to thank Benjamin Fellows and Billy Stansfield for help with manual digitization of RCHs. Katherine Johnson for digitization, feedback and comments on the thesis. Samantha Dow for feedback and comments on the thesis. Conrad Vispo for RCH data from Columbia County, NY. Chandi Witharana and Andrew Jolly-Ballantine for feedback and comments on the thesis and serving as committee members. The UConn Department of Geography for funding to attend and present my research at AAG. This work was supported by National Science Foundation grant BCS-1654462.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract:

Advanced machine learning combined with widespread, publicly available, airborne light detection and ranging (LiDAR) data has great potential to automate the extraction and classification of landforms and 17th-early 20th century land use features preserved under forest canopy throughout the northeast US landscape. Previous studies have shown that stone walls, house foundations and relict charcoal hearths (RCHs) stand out clearly in derivative LiDAR digital elevation model (DEM) products such as slope and hillshade maps, but to date, mapping has been mainly carried out by on screen digitization. In this study, a deep learning convolutional neural network (CNN) algorithm was employed to extract relict charcoal hearth features from LiDAR data. With the application of CNN algorithms on LiDAR based slope maps and edge detection rasters, the network was successfully able to identify locations likely to be RCHs. The model results were further refined using object-based segmentation and image analysis methods, and compared with a dataset of manual digitized RCHs in 5 test area that cover a range of landscape types (steep terrain, rough terrain, lower gradient valleys and wetlands, developed, and forested). The CNN approach offers value in speed and ease and performed best when applied to forested hillslopes >10 degrees and areas lacking an overprint of mid to late 20th century development. Overall, the results of this study offers a unique insight into mapping past land use activity anywhere LiDAR data exists and the landscape experienced similar land use change during the 17th-20th centuries.

Keywords: Remote sensing, Machine Learning, Neural Networks, Charcoal Hearths

1. Introduction

1.1 Introduction

Publicly available, widespread airborne light detection and ranging (LiDAR) data has become an essential tool for landscape studies by granting unique insight into the land surface morphology that lies beneath the densely forested landscape of New England (Johnson et al., 2015; Lasaponara and Masini, 2009; Witharana et al., 2018). Across the northeastern United States, LiDAR datasets have become more and more prevalent at 1 m resolution or better (MassGIS, CTECO, NYGIS, VGCI, Open Topography). Coinciding with the availability of these datasets, an increasing number of visualization techniques have been used to identify landforms and archeological features across different landscapes (Bennett et al., 2012; Challis et al., 2011; Hesse, 2010; Kokalj et al., 2011; McCoy et al., 2011). Despite this wide range of visualization techniques and advancements, feature extraction is most commonly accomplished through manual identification and digitization (eg., Johnson and Ouimet, 2016). Manual digitized datasets are valuable inputs to train and develop further automated extraction methods for location and identification (Cowley, 2012; Witharana et al., 2018).

This paper presents a deep learning convolutional neural network (CNN) for the automated extraction of relict charcoal hearths (RCHs) from LiDAR data and derivative raster maps in northeastern US. The northeastern US is characterized by history of deforestation for charcoal production, agriculture and pastoral activities from the 17th-early 20th century, followed by abandonment and reforestation (Hall et al., 2003; Foster et al., 1998, 2003, 2008; Johnson and Ouimet, In Review). LiDAR penetrates the modern, dense forest canopy of the region and reveals many historic land use features, including stone walls, foundations and relict charcoal hearth (RCHs). Remote sensing

and geospatial analysis hold great potential for studying these historic features. Previous studies have been focused on identification of stone walls, foundations, and dams in LiDAR data (Johnson et al., 2015; Johnson and Ouimet, 2014; 2016), as well as impact of RCHs on soils (Raab et al., 2017; Hirsch et al., 2017; Flannery et al., 2019) and RCHs' connection to historic deforestation (Stansfield and Ouimet, 2019; Johnson and Ouimet, In Review).

RCHs stand out from the background topography as flat elliptical features with steep edges (Figure 1). In previous studies, digitization of RCHs has been done by building a grid of the area in question, and carefully scrolling through and manually identifying and digitizing RCHs by placing a point in the center of the circular or ovoid features (Johnson et al., 2015; Johnson and Ouimet, In Review). Other methods which have been used include template-based point cloud analysis and object-based image analysis (OBIA) (Schneider et al., 2015; Witharana et al., 2018). Template-based analysis was successfully applied to LiDAR point clouds in Germany, where features are more regularly shaped and are located in similar flat areas (Schneider et al., 2015). The success was largely based on the features being the primary differentiation from the background, rather than the more complex landscape of New England. OBIA, meanwhile, has had some success when applied to small scale areas of high RCH concentration in the Connecticut within southern New England (Witharana et al., 2018). OBIA "segments" images by combining like pixels into image objects and using qualities of these objects to divide objects into classes. Segmentations rely on relationships such as sharp lines where objects switch from high to low slopes or neighboring pixels with vastly different values. The OBIA rule sets identify factors of these image objects to

identify other objects which share similar characteristics and classify those into similar classes (Witharana et al., 2018). This is similar to the deep learning methods that were used in this study, with the differentiation being that the user has to manually identify factors which identify the RCHs in OBIA, while deep learning methods use computer analysis to identify patterns. The overall objective of this study is to create an algorithm which can automate the extraction of RCHs to an equal or higher accuracy compared to manual digitization. Specifically, this study will discuss: 1) progress towards the goal of creating a convolutional neural network which identifies RCHs across different terrain; 2) the transferability of the network across different landscapes; and 3) implications of the network's development and future work necessary to identify RCHs in locations where the network falls short.

1.2 Historical Context and Morphology of RCHs

RCHs and charcoal production in general provides a unique insight into the land use and economic history of New England. Charcoal was primarily used from the mid-1700's through the early 1900's in the northeastern US to produce the fuel needed to process mined iron (Straka, 2004). RCH distribution varies throughout the region from high density, typically near large scale furnaces, to much lower density and sporadic, where charcoal was produced for local subsistence trade by farmers and foresters. LiDAR based mapping of RCHs show the regional distribution of charcoal production and how prevalent they were throughout Connecticut (Figure 2). This distribution of charcoal production and level of detail is not unavailable through historic maps and records. In Connecticut, large scale charcoal production took place in the northwestern portion of the state beginning in 1760 and ending in 1920 in support of the Salisbury

Iron District (Johnson and Ouimet, In review). In this portion of the state, 20 furnaces which ran for five or more years, with production peaking with 16 different furnaces running at once in 1858 (Johnson and Ouimet, In review; Gordon, 2001; Gordon and Raber, 2000; Kirby, 2011; 1998). Around 1900, temporary RCHs built on hillslopes were replaced by stone and metal kilns and the introduction of easily available anthracite coal replaced the demand for local charcoal production (Barger, 2013; Journal of the United States Association of Charcoal Iron Workers, 1885). Neighboring states of New York and Vermont had a similar history of iron industry and charcoal production (Rolando, 1992; Gordon, 2001).

A typical furnace would require 600 acres annually to run the iron production. With the 20 different furnaces this means that loosely 12,000 acres were in rotation (Straka, 2014). RCHs were created to provide flat areas where colliers would stack cut logs and use concentrated burns to harvest large amounts of charcoal in a single burn (Figure 1A). Colliers would grade out the soil from an area on the slope and stack piles of logs from 4 to 6 meters tall for each use of the hearth. Remains of these hearths are circular, flat features ranging from 7 – 12 m diameter in size, with an edge allowing for the charcoal to be loaded by colliers onto horse drawn carriages. The hearths were primarily built in densely forested areas and were harvested every 20 to 40 years, leading to a multilayered pattern of charcoal and soil remaining today in the hearths (Hirsch et al., 2017; Schwarz, 1907; Winer, 1955). These platforms leave topographic anomalies easily identified in the LiDAR based on the difference of the top of the flat hearths from the background slopes (Figure 1).

Over 24,500 RCHs have been identified across eastern New York, Connecticut and western Massachusetts (Johnson and Ouimet, In Review; Figure 2). The background slope that RCHs are found on largely determines the morphology of the hearths. RCHs in LiDAR slope maps exhibit two distinctive styles – 1) a flat platform inset within a steeper surrounding hillslopes and 2) a flat or raised platform surrounded by a circular ditch on lower slopes or flat terrain. RCHs built on slopes are carved out from the background, range from oval to circular, and are typically 7-12 meters in diameter. At the highest background slopes, these RCHs have a more teardrop shape. This type of RCH is easily identified by the differentiation of the flat RCH platform from the background slope. In areas where the background slope is low or relatively flat, RCHs are generally bigger (8-15 m in diameter) and are raised with a ring or ditch partially or completely surrounding the feature. These are almost completely circular with high edges but are less identifiable on slope maps because the lower slopes are often less distinguishable from the background slope (Figure 3). Despite the differences, the prominent edges combined with the circular nature and consistent range of sizes define the RCHs. In the study area, the greatest number of RCHs are found on slopes of 15 degrees or higher (Figure 4) (Johnson and Ouimet, in review).

2. Materials and Methods

2.1 Building and Training of the Neural Network

2.1.a. Background and Workflow

The deep learning method applied in this study was a convolutional neural network (CNN), which is a specific neural network designed to recognize patterns in an image while minimizing processing power based on random initial weights of different features.

Neural networks function by analyzing patterns across multiple layers of information and comparing samples previously identified as “targets” compared to samples generated on background or “unclassified” areas. Pixels are grouped by patterns of contrast in a hidden layer which groups like patterns (Figure 5). Convolutional networks lower the processing power by random initial weights for the different correlation patterns. Samples include information from any number of image layers and based upon consistent patterns hidden networks are built identifying consistent patterns across inputs.

The neural network that was used for feature extraction in this study was included within the eCognition software package (Trimble, Munich, Germany). Within eCognition there are pre-constructed algorithms which allow manual inputs of the sample patch size, image layers, number of hidden layers, number of feature maps, kernel size, and max pooling. The output from the CNN returns heatmap with values for individual pixels ranging from 0 – 1. These values represent the likelihood of each pixel to be part of an target (i.e., an RCH), where 0 means unlikely and 1 means highly likely. The network requires a training region, a validation region and a test region (Figure 6). The network is trained on known samples of the target feature within the training region, and once trained, it is applied to the validation area. Outputs generated from validation results had recall, precision, and F1 scores calculated compared to a manually digitized dataset. Inputs are then adjusted in order to increase true positives while minimizing false positives and negatives. Training the network is an iterative process and outputs were produced at various learning rates with varying amounts of training steps. The learning rate and training steps for the final network was chosen based on the highest F1 score

for the result. Once the combination of learning rates and training steps are chosen, the network is applied to unknown test regions. The resulting outputs indicate the transferability and the effectiveness across different land cover types and topographic landforms across the study region.

The neural network that was used for feature extraction was completed in eCognition and ArcGIS using the following steps: 1. Create rasters based off of the LiDAR DEM; 2. Input the slope and Canny edge rasters as well as manually digitized RCH samples into eCognition; 3. Divide the region into the training and testing regions; 4. Segment the training region identifying the centroids of RCHs as the target image objects; 5. Create the neural network, generate samples from RCH objects and unclassified objects and train the network with the sample patches; 6. Apply the trained algorithm to test regions; 7. Post processing on the output likelihood raster and cross-validation with manual digitization results (Figure 7).

2.1.b. Inputs and Network Training

The input datasets are created from airborne LiDAR datasets retrieved from Connecticut Environmental Conditions Online (CTECO, 2016; CRCoG, 2016) and NOAA Digital Coast (NOAA, 2017) in the form of LAS tiles. CTECO publishes the data in digital elevation models (DEMs) with a 1 m pixel resolution. The point clouds (.las files) have a average point spacing between 0.7 m to 1 m and are based on “2-Ground” processing. Manual digitization of RCHs has been done primarily using hillshade and slope rasters generated from the LiDAR DEMs in ArcGIS, and the training area was chosen from an area of high RCH concentration (> 80 RCHs/km) (Figure X). The entire state of Connecticut has been mapped at coarse scale to identify the presence/absence

of RCHs. in this coarse scale mapping, grid squares are approximately 1 km^2 and are given a value of 1 if there is greater than 1 RCH in the area or 0 if no RCHs are present. This gives context to where to apply the network within Connecticut. Using a single kilometer grid cell with known RCHs will allow the network to minimize false positives while only using areas where there are known to be true positives.

In this analysis, there were two input layers for the CNN, the first being a standard slope map and the second being a Canny edge detection map (Figure 8). The slope was created by calculating the maximum rate of change from the elevation of a cell to that of its neighbors. This raster can be calculated as either degrees or percent rise. Degree measurements range from 0 (completely horizontal) to 90 degrees (completely vertical) whereas percent rise is calculated by rise over run multiplied by 100. In the case of building a deep learning network, a degree measurement is used and an edge detection algorithm is built to identify the edges around the flat zones on the top of the RCH. The edge detection used is a Canny edge algorithm introduced by John Canny in 1986 (Canny, 1986). Canny edge detection is broken down into five parts: first, a Gaussian filter is applied to smooth out noise from the slope map, then the intensity gradient is calculated in order to determine the horizontal and vertical direction of the edges. Following that, an edge thinning technique called “non-maximum suppression” is applied to identify the largest edge and remove additional responses to the filter. A double threshold is used to highlight edges over the high threshold and repress the values below the lower threshold value in order to smooth out the unimportant edges. Finally, weak edges are removed by a form of blob analysis (an analysis to identify connected pixels) unless they share a strong edge pixel in the same cluster. The

algorithm used in this study was included in the eCognition software package which performs the analysis and outputs a raster highlighting the edges.

Slope rasters were produced from the DEMs within ArcGIS (ESRI, 2019) and were input into eCognition as an 8 bit dataset ranging from 0 – 90 degrees. In eCognition, a Canny edge raster was created from the slope dataset with values ranging from 0 – 1 (Trimble, 2019). Both Canny edge and slope maps were normalized to a 0 – 255 range in order to ensure both inputs were valued equally within the neural network. Within the training area, 8,000 samples were generated as 40 pixel squares based on the RCH centroids. The 40 pixel sample size was chosen based on the size of the RCHs. With the LiDAR being 1 m per pixel and an average 7-12 m diameter RCH means the minimum size of samples must be at least 12x12 pixels, additionally 3-5 m were included for the ridge. With an allowance for larger RCHs, 20 m was left surrounding the center pixel with 10 m allowed on each side for the network to recognize the background slope patterns. Samples needed to include enough area to encapsulate the flat platform of the RCH and the steep edges, which provides the pattern to differentiate the entire sample from background slopes which do not include the target. Alongside these samples of the target class, an additional 4,000 samples were generated from unclassified areas to highlight what makes the RCH samples unique. From there a 2-layer convolutional network was created utilizing a single hidden network with a max pooling value of 40 and a kernel size of 13. It was trained using a learning rate of .0009 with 100,000 training steps of 50 samples per batch. The learning rate was chosen after networks were tested ranging from 0.015, 0.01. 0.0009, and 0.0006 were tested. The training rate was decided on after analyzing network results

with 10,000; 20,000; 30,000; 50,000; 100,000; and 500,000 steps. F1 scores increased until 100,000 steps, where results became too specialized to the input and when applied to validation and test areas the results decreased in accuracy.

2.1.c. Training area

The area used to develop and train the neural network was a 2.5 x 2.2 km area in Cornwall, Connecticut which contains 208 RCHs (Figure 9). Of these RCHs, 169 were easily identifiable with a slope above 15 degrees, and 39 were found at a slope below 15 degrees and were determined as difficult to identify. By only using the higher sloped samples the network was able to be trained in an area where it would have high success. By ignoring RCHs that were difficult to distinguish from the background slope and the RCHs within the rough areas of terrain, the network was able to have a much higher identification rate at high slope areas, which composed over 90% of the dataset (see Figure 4). Using samples from the rough terrain was found to increase false positives within all areas of rough terrain, even if no RCHs were present. The vast majority of RCHs within test and training regions were located in areas with a background slope greater than 15^{o}. Therefore, the network was designed around identifying the highest percentage of the more common type of RCH while ignoring the sub 15^{o} RCHs and only used samples generated from higher sloped RCHs. This increased the overall accuracy and decreased the false positives in the final dataset.

2.2 Processing the Validation Region and Analysis of Inputs

The validation area (Figure 10) was an approximately 0.5 km^2 area located in Litchfield, CT that was known to contain a high concentration of RCHs. This area has 58

known RCHs, and of these, 3 are located in areas with below 15^{o} slope so they would be expected to be harder for the network to identify. This area contains some roads which are an expected false positive and the southeastern corner of the area has a river running through which is another expected false positive. This location was chosen as a validation area because it includes features that are expected errors associated with the network, and for its high concentration of RCHs with background slopes that would be expected to be identified.

The initial output of the CNN was a heatmap of the entire validation area with values ranging from 0 – 1, representing the likelihood each pixel being included as part of an RCH (Figure 10). The output heatmap was smoothed in eCognition using a Gaussian filter in order to remove the singular pixels from the output. The Gaussian filter is a form of a low pass filter which used a 5x5 kernel to reduce image noise and therefore remove singular high value pixels from the results. Gaussian smoothing greatly reduces average pixel values, leaving the majority of unidentified pixels at or close to the 0.0 mark (Figure 11). Once the heatmap was smoothed, a multi-threshold segmentation algorithm was applied to the smoothed heatmap in eCognition. This assigned the image class of Target to clustered areas of 3 or more pixels with a likelihood of 0.9 or higher. The image objects were converted into vector format, producing polygon shapefiles that could be brought into ArcGIS for further analysis.

When these polygons were brought into ArcGIS the Feature to Point tool was used to generate a point shapefile of the centroids of these polygons. From there, an 8 m buffer was generated based on each centroid and compared to previous datasets for validation. The validation area RCHs were already known, and these centroids also had

an 8 m buffer applied to generate the outline of RCHs. The two shapefiles were intersected and overlapping buffers were considered true positives, buffers from the manual dataset that were missed were classified as false negatives; RCHs identified by the network that did not overlap with manual identification were considered false positives. Validation analysis used precision and recall measures to analyze the results of the network. Recall is the ratio of positives identified by the network to the actual number of true positives in the region and is measured by the equation: true positives/(true positives + false negatives). Precision, on the other hand, is the ratio of true positives compared to all positives identified by the network and is measured by the equation: true positives/ (true positives + false positives). The final score for analysis was the F1 score (2x (recall x precision)/(recall + precision)) which averages the two scores to create a singular value for accuracy.

In some cases, the output heatmaps will still identify stone walls, roads, houses, and stream beds as they share morphological features of low slopes with adjacent high edge factors. In order to remove the signatures of these features, a mask of the test area was created removing the waterbodies (based on waterbodies and hydrology polygons obtained from CT DEEP (CTDEEP, 2008) and the impervious surfaces layer from CTECO (Quantum Spatial, 2017)). These buffers were removed from a shapefile of the training area using the Erase tool (ESRI, 2019), and the pixels were masked out of the dataset by the Erase by Mask tool (ESRI, 2019).These buffers removed a large number of false positives from the dataset and helped to create high quality results without prior vetting of the test regions. After all post processing is completed, approximately 27% of false positives are removed, however, the waterbodies layer that

was used does not align perfectly with the LiDAR based background dataset and shows limited actual effect on the analysis. Within the test regions used for the final step of CNN analysis, the hydrology lines were reshaped to better fit the LiDAR signature of waterbodies (Figure 12). These reshaped stream and river beds were used on the analysis of the validation and test regions, but were not used in the overall accuracy calculation in order to the emphasize the way the algorithm could be applied to a large region.

As RCHs themselves impact the slope with the flat surface, two buffers were created from the centroid: a 10 m buffer and a 14 m buffer. The 10 m buffer was then Erased from the 14 m buffer leaving a ring surrounding the RCH buffer. Zonal statistics was applied to the background slope to find the mean value of the slope within that ring. Box and whisker charts were made to identify where the high/low slope cut off was the most effective at dividing the dataset into different and the 20 degree mark was identified as the transition between higher and low slope values based on the results shown in Figure 13.

2.3 Results of Test Regions

The full test area is a 24.6 x 11.1 km rectangular surrounding the validation area in Litchfield County, in northwestern Connecticut. The area spans across the towns of Canaan, Norfolk, Winchester, Barkhamsted, New Hartford, Torrington, Goshen and Cornwall. This area contained topographical variation typical of the New England landscape, including developed areas, and waterbodies within the region. The test region was identified from the RCH presence/absence map of Connecticut (see Figure 2) and contained over 4,000 RCHs. Northwestern Connecticut was chosen because of

the high concentration of RCHs within the region as well as publicly available datasets for post processing of waterbodies, hydrology, and impervious surfaces.

From this large test region, six small scale examples of the different successes and failures of the network were selected and analyzed individually (Test areas 1-6, see Figure 14, Table 1). These areas highlight where the network has the highest identification rates and the regions where there are the greatest number of false positives or false negatives. Test areas 1, 2 and 3 exemplify areas where there is a high concentration of sloped RCHs where the network is expected to perform well. False positives in this region are likely from the impervious surfaces more so than any natural topographic variation. This section of the test region is a prime example of a region where the network should perform well, as most RCHs occur on sloped areas and there is limited topographic variation once impervious surfaces and water bodies are processed out. Test area 3 has a sloped region but also shows what happens when low slope is intermixed with high slopes. This would be expected to have low false positives, but a lower recall rate as some RCHs are found on low slopes. Test areas 4, 5 and 6 represent the areas where the network is expected to perform very poorly. Test region 4 is a section of the landscape with primarily low slope that was identified as part of the region which has multiple RCHs present (Figure 19). Based on the inputs to the training of the network, this is an example of an area where RCHs would be classified as missed true positives. Test areas 4 and 5 on the other hand exemplify difficult regions based on large amounts of false positives. Test area 4 is a developed section of the test region where there are likely no RCHs present. This is an example of a test site where there would likely be a large amount of features which could generate false positives

despite there being no true positives in the test. This exemplifies the importance of knowing where to apply the network, rather than applying without consideration of the regions. Test region 5 is another example of where there is likely a large amount of false positives. This region is sloped, but it is characterized by rough terrain. Within the roughness, false positives arise from the local relief of the bumpiness of the landscape.

Initial results from the algorithm across the entire test region returned a recall value of 0.82 (correctly identifying 82% of RCHs) (Table 2). However, the recall further increased to 0.87 when inspecting samples exclusively on slopes above 15^{o} and even further when discounting values below 20^{o}, increasing the recall score for the full area to to 0.95. While recall values were very promising, precision was highly lacking at 0.22. This is because there were almost four times the amount of the false positives compared to true positives overall (Figure 15). The end result were F1 scores of 0.34 for the entire region, increasing to 0.35 and 0.36 respectively as the lower slope RCHs were removed.

In comparison to the results for the entire region, Table 3 shows the results of the individually analyzed test regions. Results were similar to the expected outcomes, with test areas 1-3 and area 6 having a 0.8 or greater recall, while areas 4 and 5 have very low scores. Large numbers of false positives occurred throughout all datasets with the exception of test areas 1, 3, and 4. Overall trends were as anticipated, and the results are much more promising than the blind run over the entire test region.

Test areas that contained slopes of 15^{o} or greater had similar recall values and significantly higher precision and F1 scores than the blind application. Test area 1 correctly identified 44/49 hearths, however the region also contained 40 false positives.

As seen in Figure 16, the majority of false positives come from impervious surfaces or stream beds. Of the 40 false positives, only 12 were on a slope outside of the developed region of the test area. Of the missed RCHs, 2/5 were not well developed with raised surfaces running through the normally flat top of the RCH which makes them difficult to pick up by the algorithm. Test area 2 (Figure 17) successfully identified 89% of the RCHs within the region but is an example of where the false positives heavily outweigh the true positives. With 22 true positives and 74 false positives, there are almost 340% more false positives than true positives in the test area. The missed RCHs occur primarily on sub-15^{o} slopes, however the network does not perform particularly better at higher slopes. There are false positives in the rough terrain in the northeastern corner and some intermixed with successful areas. A more accurate LiDAR hydrology layer would remove at least 7 more false positives and there are a large number of false positives within the developed region. Test area 3 (Figure 18) demonstrates an area where there is bimodal topography. On the sloped regions, the identification rate is close to 90%, while on near 0^{o} slopes, the identification rate is closer to 50%. Overall, 74/79 high slope RCHs were identified, and including the low sloped regions, there is still an over 0.8 precision score. There are limited developed regions and limited hydrology within this test area and this is reflected in the false positives. Of the 7 false positives, 2 are from the two water sources in the area, and 2 more border roads. This area represents the effectiveness of the network in areas without water or developed areas.

Regions with lower slopes, large portions of developed land and rough terrain explain the vast majority of the false positives and errors explaining the low precision

value in the blind test results. Test area 4 (Figure 19) is an area where the network does not perform well. This is an entirely low sloped region (all RCHs within the region have a sub 15^{o} background slope) and contains only 1 successfully identified RCH out of 12. There are also 8 false positives within the region, leading to an 0.08 recall and 0.11 precision score. This test area represents the need for context when applying the network. Test Area 5 (Figure 20) is an example of when the network is applied to an entirely developed region with no RCHs present. There are 26 false positives from a combination of stone walls, impervious surfaces not included in the layer used for post processing, and waterbodies that were not removed. Applying the network to an area such as the test region which includes several exclusively developed regions, demonstrates why the overall precision of the algorithm is so low. In theory, an area like this would be the least effective usage of the network. Test area 6 (Figure 21) is an example of where the network may be expected to perform well but does not. There is only 1 missed RCH within the test region, but there are 55 identified false positives. This area has a very high recall rate (0.91) but a precision of only 0.15 with an overall F1 score of 0.26. The roughness of the terrain in this area creates a large number of false positives as many of the ridges and “bumps” in the topography cause features with very similar edges and flat tops to the RCHs.

3. Discussion

The neural network has shown very promising results overall when applied to the regions where RCHs can be expected. There was an overall recall score of 0.817, where higher results were expected after cross validation with digitized RCHs. There was a significantly higher lower identification rate (the rate drops by ~70%) on hearths

with a background slope of less than 15^{o} (Table 2). This was based on the input raster datasets better highlighting the morphological anomalies of the high sloped RCHs. In both cases, the Canny edge algorithm does well identifying the edges of RCHs, however at low slope, the slope raster is less useful, as there is the RCH is less differentiated from the background slope. In addition, a significant amount (~ 7% (67/869)) of the RCHs with sub 20^{o} background were removed when a Gaussian filter was applied to smooth out the results. This was intended to remove the pixels which were incorrectly identified in single pixels or small clusters; however, this did remove areas with a very low amount of high likelihood pixels.

Test areas 1,2 and 3 showed the most ideal topographic landscapes for the neural network to be applied. These regions were sloped throughout the test region and while they contained stream beds, waterbodies and impervious surfaces they were mostly able to be masked out. Within these test areas there was an over 0.8 recall rate, and of the missed identifications, several required landscape context to fully identify the missed samples. While these test areas show that mixed landscapes are possible, test area 2 however, highlights the difficulty of blindly applying the model to all areas with the higher slopes. Within the sloped samples, there are a large number of false positives from minor relief changes and manmade features. Manmade features cause variation in the landscape similar to that of RCHs where there are flat roofs or pools that share similar morphological features.

It is important to contextualize the results of the neural network with the fact that the dataset used for identifying true or false positives was based on manual digitization. Study area digitization was completed using LiDAR only and no on sight investigation

was used in the identification process or ground truthing. It is possible, therefore, that some true RCHs were identified by the neural network that the manual dataset missed. This would lower false positive counts and increase the precision of the results significantly. This concern is especially magnified in test area 6 where the rough terrain has many examples of false positives of similar morphology. While current analysis identifies only 11 RCHs in the region, some of the 55 false positives could be found to be true RCHs upon testing the soils on site. Is this manner, the neural network results should be considered an important tool for mapping new RCH locations, and areas with a high number of false positives reported in this study highlight regions that could be investigated for RCH presence in the future, hence the there is no reason to rule out false positives as potentially true RCHs.

When the network was applied to the entire large test region (25 x 11 km) there was no landscape context used to ensure the tested areas were suitable to RCHs. This meant that areas of heavily development and near zero slope areas were both included within the major test area. Therefore the addition of previously digitized datasets was required to increase the accuracy of results (clearly seen in the developed region in Figure 20). The layers chosen for post processing were an impervious surfaces layer (Quantum Spatial, 2017) and the waterbody and hydrology polygons (CT DEEP, 2008). The impervious surface layer shown in Figure 12 was created based on LiDAR images and effectively removes signatures from buildings. In areas where there is no available impervious cover shapefile, the “developed” class of high resolution land cover rasters (accessed from CTCLEAR) proved to be equally effective. This figure also demonstrates the issue with the other input layers. False positives are caused where

water bodies, especially those of a smaller, circular nature, are found within the analysis region as these features have flat areas and often have steep edges. As previously mentioned, the impervious surface layer serves to remove approximately 21.9% (3,790/17,288) of false positives overall, while the waterbodies and hydrology layers (CT DEEP, 2005; 2008) only remove 6.1% (1,063/17,288). Of the false positives remaining in the dataset, a large portion are likely waterbodies which are not removed by the dataset not covering all the smaller waterbodies as well. Additionally, stream beds cause issues with identification as building a buffer from hydrology lines is ineffective based on the variable nature of the stream bed. Because the width of the stream varies, using a single uniform buffer will cause issues both with under and overcompensation for various parts of the stream bed. Ideally, LiDAR based datasets of stream/river beds, waterbodies and other features would be available to the precision of the impervious surface layer. This would serve to remove the vast majority of false positives, even then, it would likely not be able to remove them all as there are many variations of 5 – 20 m within the landscape that can cause similar patterns to RCHs. Without the context of a landscape, even features such as pools or circular driveways can create false positives within the datasets which make the application of the network without studying an area ineffective.

Variation within the hearths cause an issue of its own. While differences can be highlighted between the low and high sloped areas by characterizing background areas, unique use patterns also affect the final morphology. Duration and amount of use serve to accentuate the individual patterns of hearths. The more times a hearth was used, the more distinctive the remaining features are, while hearths used only a single time are

much less differentiable from background slope. With the differences between hearths, combined with inputs, manual and neural network digitization do not always agree on results. However, manual digitization errors are one of the most harmful to the training and accuracy of the network. If RCHs appear within the unclassified samples of the network, it decreases the ability of the network to identify the difference between RCH samples and unclassified “background” slope samples. The network works by identifying patterns which appear in the target samples which are anomalous from the background, where missed manual inputs cause the program to be unable to determine which patterns are unique to the RCH feature. Other issues possible from manual digitization is it can be difficult to place a point in the direct centroid of a sample. This causes the network sample areas to include large sections of non – RCHs and makes it difficult for the network to identify consistent patterns as pieces of the RCH will not be included.

The accuracy of the manual digitization was assessed by comparing the results from RCHs digitized off 2012 LiDAR to systematically digitized results from the updated 2016 LiDAR (CTECO, 2016; CRCoG 2012; 2016). Underpinning the accuracy of any results is the input data from the LiDAR. Ground and last return points that generate the DEM inputs, and which all derivatives are based on, are heavily influenced by forest cover and underbrush. If the initial ground Earth is inaccurate then all derivatives produced later in the process will not have an accurate picture of the area. Water features in particular cause issues within the landscape as, these features are removed and smoothed during airborne LiDAR analysis. This creates a false smoothed slope with steep edges where the land is no longer being removed by smoothing. These high

edges are often selected as false positives by the RCH network. Within the area alone there is a significant difference between the results from different users, especially when different years of LiDAR data were used. From this comparison, the accuracy of the algorithm needs to be highlighted as a valid alternative in both accuracy and the speed of application.

Further analysis on alternative methods to enable low slope RCHs to be recognized is required, but likely simple slope and edge detections will not be the correct inputs. Figure 16 clearly demonstrates the lack of results when applying this network. Applying modern networks to past features, especially those hidden by forest cover is challenging as aerial photography is not a valid input for help with identification. Nevertheless, further LiDAR based datasets for features such as ridge lines and stream beds are logical next steps for use in the analysis, which will only serve to increase the accuracy of the RCH detection. As more data becomes available on which areas are distinctly non-RCH, more false positives will be removed leading to better results. Overall, however, with landscape context and informed decisions on where to apply the network, results are highly accurate. The over 90% identification on high slope areas across several test sites and a large variety in landscapes is very promising. Applying the RCH identification network within known areas will increase the speed and accuracy of the digitization process and further unlock the distinctive past land use history of New England.

4. **Conclusion**

This study found success in the application of deep learning algorithms to automate the extraction of features from LiDAR based digital terrain models in selected

regions. This provides a viable alternative to manual digitization in areas of greater than 15 degree slopes and the basis of a network in order to identify low slope samples. Results were promising in that there were similar recall values between manual digitized outputs and automated extraction outputs across multiple test regions. In the cases that post processing must be done to samples, a baseline dataset will still provide valuable insight into areas which are otherwise still forested. Sloped, forested regions have high identification rates and low numbers of false positives, especially when compared to the developed regions. LiDAR based mapping provides valuable information about historic land use features without the large investment of on-the-ground field-based surveys. Feature extracting algorithms also rapidly provide datasets to users who do not have the prior knowledge of ArcGIS and eCognition toolsets to manually identify these features. This approach offers value in speed, accuracy and ease as well as a unique insight into the extent and spatial distribution the past land use activities in Connecticut, as well as anywhere in the northeastern US where LiDAR data exists and the landscape experienced similar land use activity during the 17th-20th centuries.

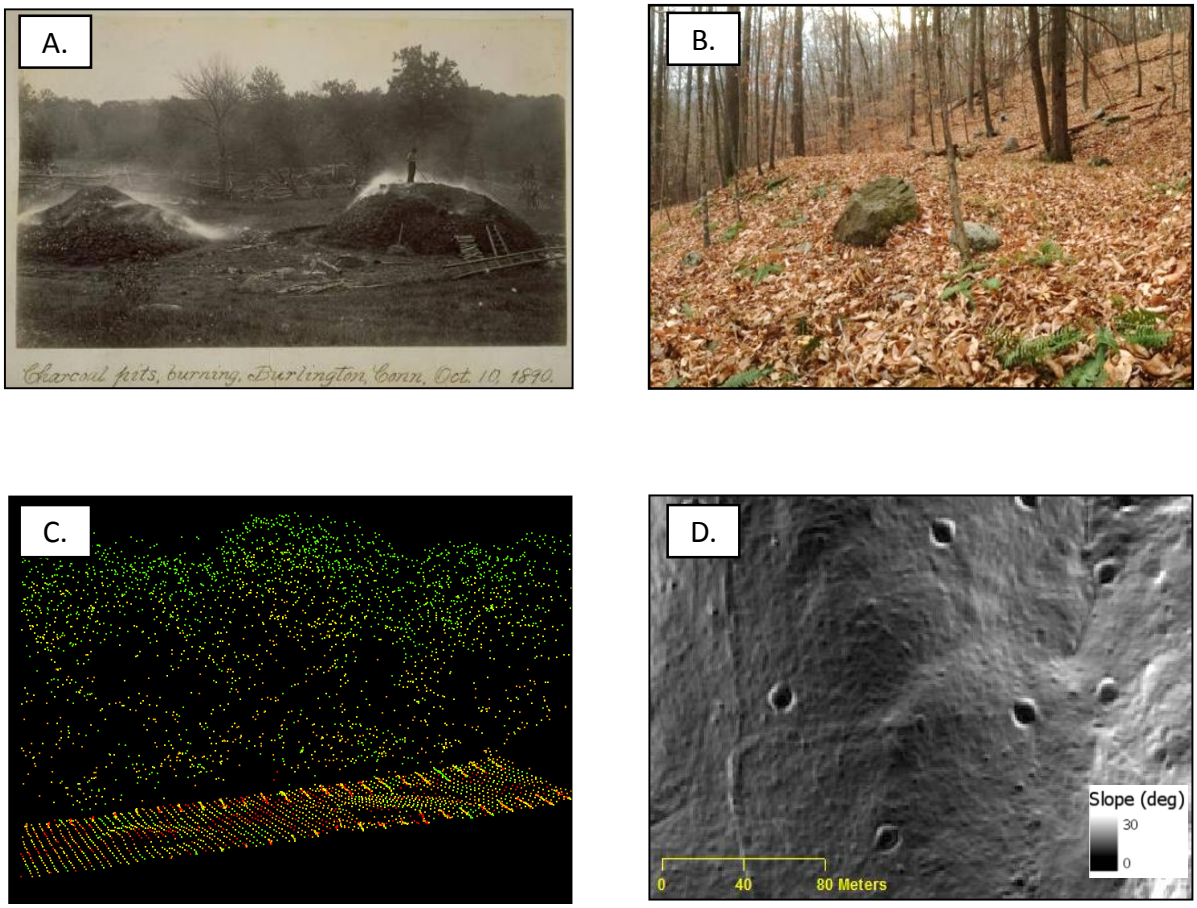


Figure 1: The life cycle of an RCH. A. is a hearth in use in 1890 B. is the remnant of hearth in the landscape today C. shows 2 RCHs in a LiDAR point cloud and D. shows RCH anomalies present in a slope map of a landscape. Figure adapted from Johnson et al., 2015. Images from Collection of the Cornwall Historical Society, Cornwall, CT. Used with permission.

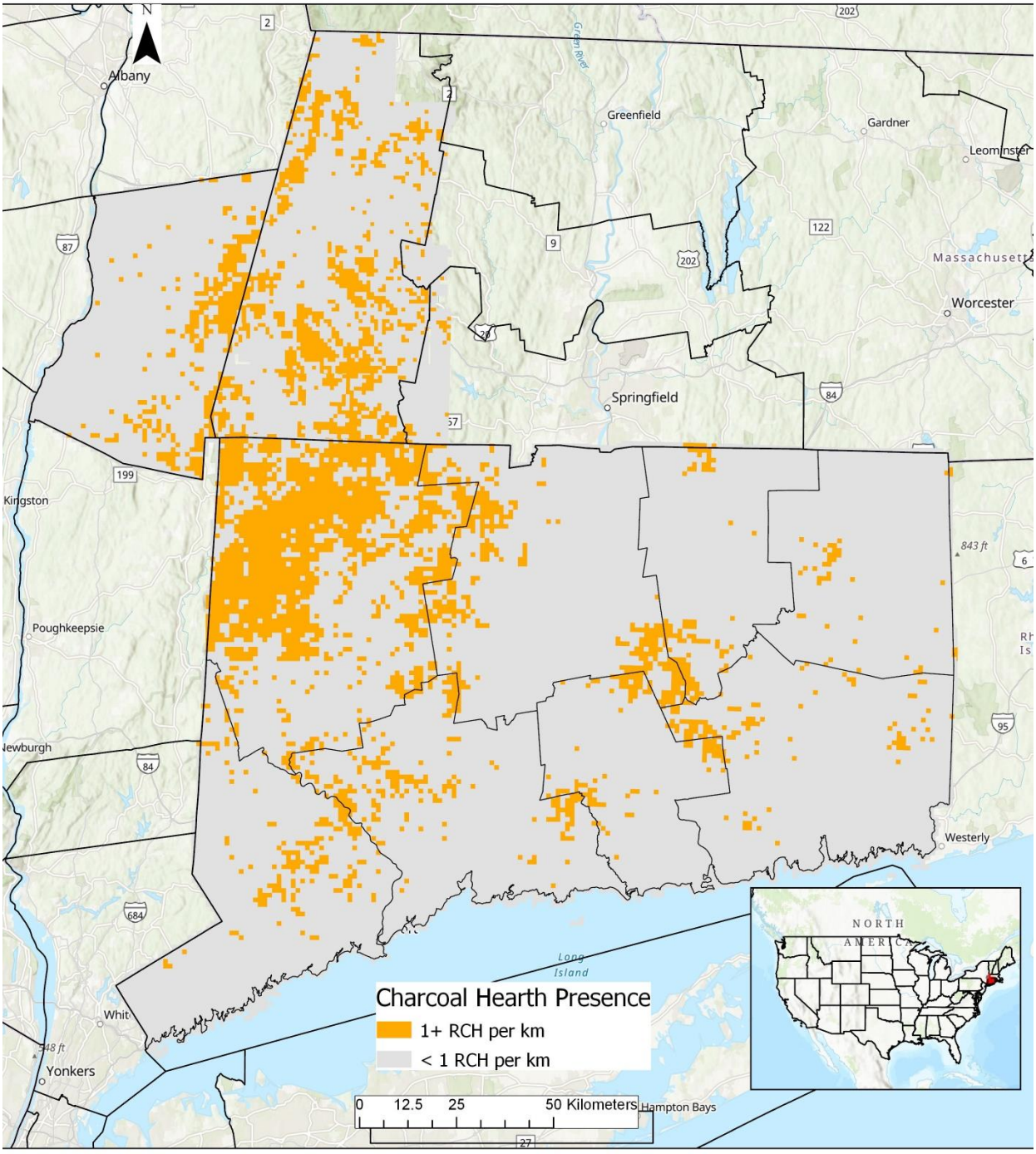


Figure 2: Presence and absence of relict charcoal hearths (RCHs) across Connecticut. There is a much higher concentration in the Northwestern corner of the state.

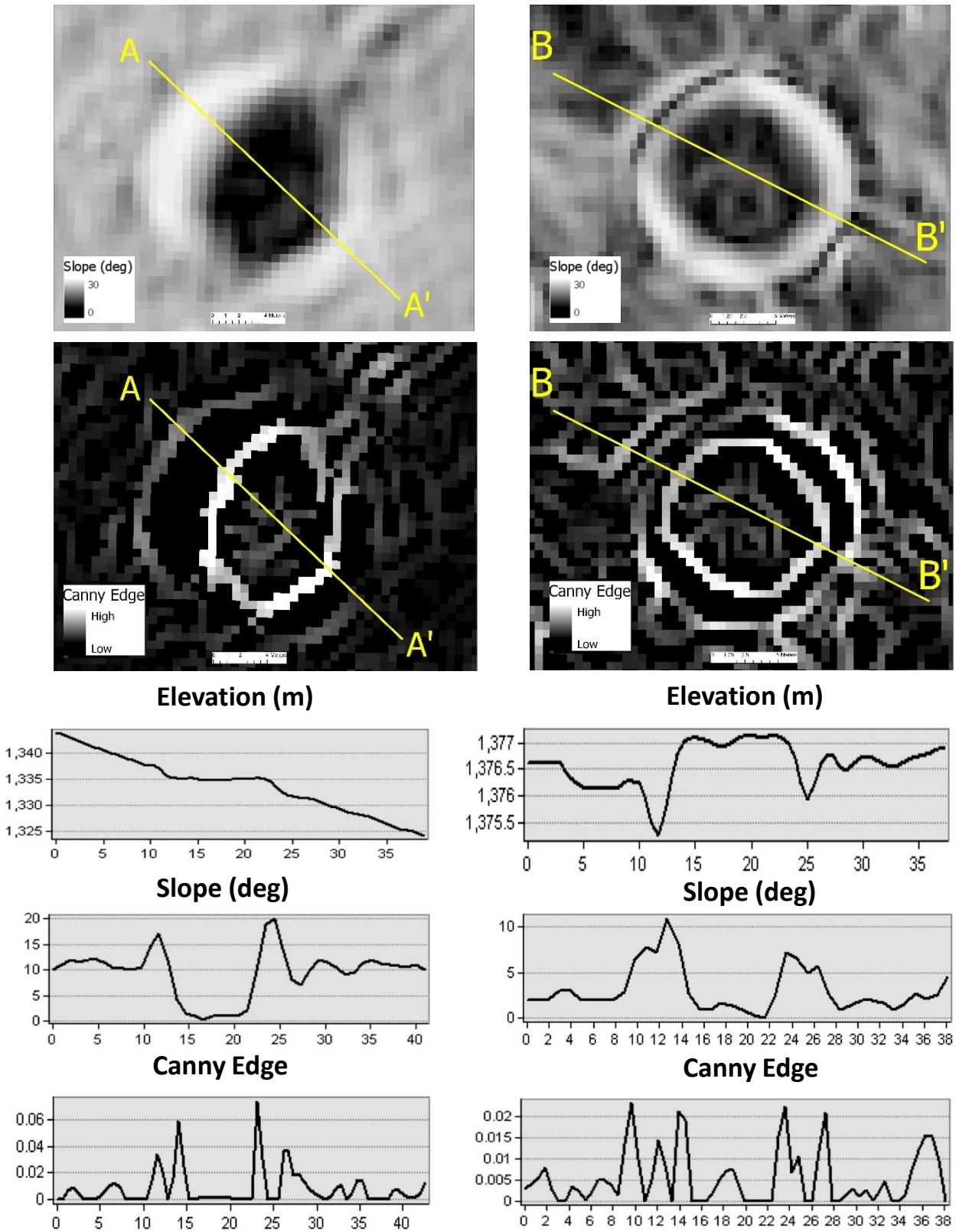


Figure 3: The comparison of charcoal hearths found on sloped landscapes (A – A') and on near zero slopes (B – B'). The high slope hearth was built into the slope, while the low slope hearth is a raised platform with a ditch built around the platform.

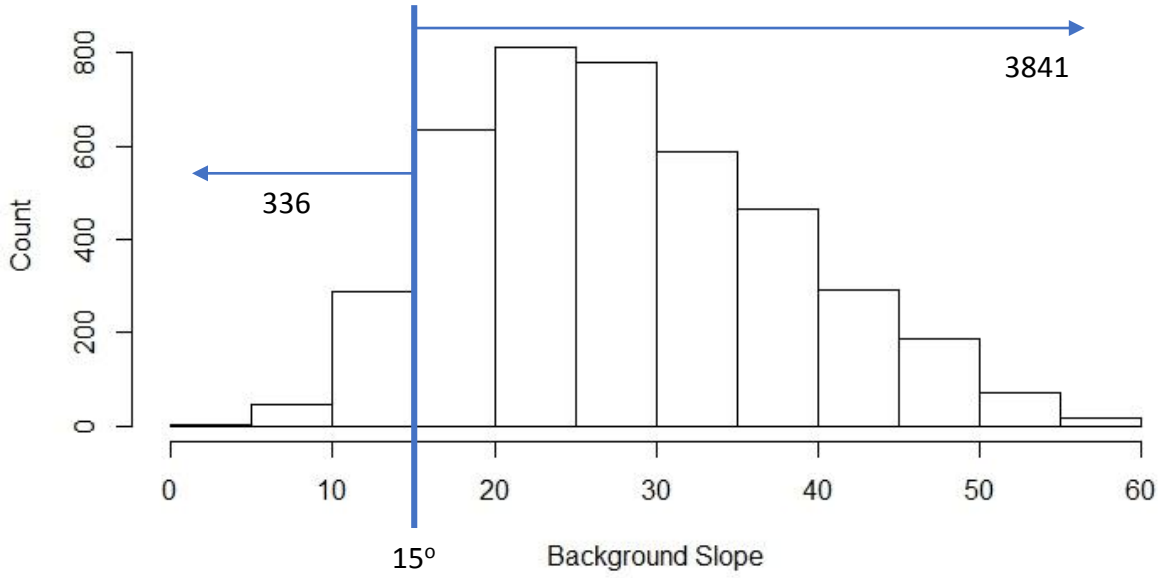


Figure 4: Histogram of hillslope gradients where RCHs occur within the test region. The vast majority of RCHs in our region were located in areas with a background slope greater than 15º with less than 10% of samples being found below that mark.

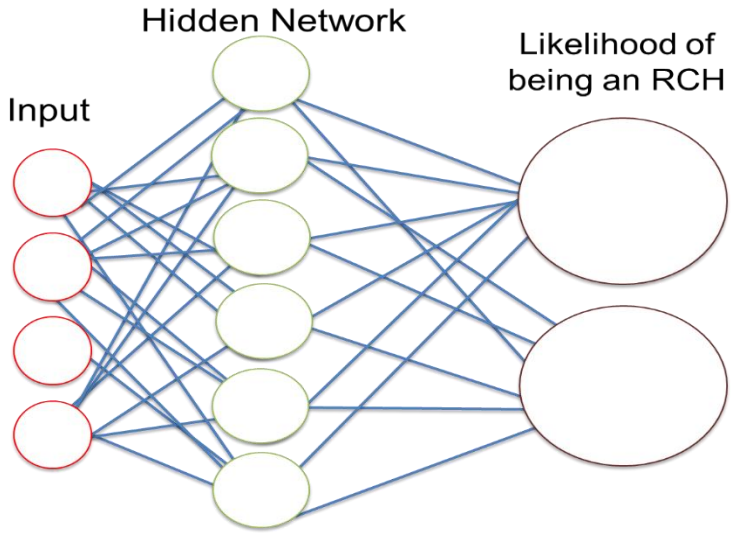


Figure 5: Convolutional neural network, pixels are grouped into objects starting with a seed pixel and identifying objects made of like pixels. The patterns of these objects are then identified and compared to the background and grouped into an output bin, in the case of this network RCH or unclassified. Neural network workflow is also given.

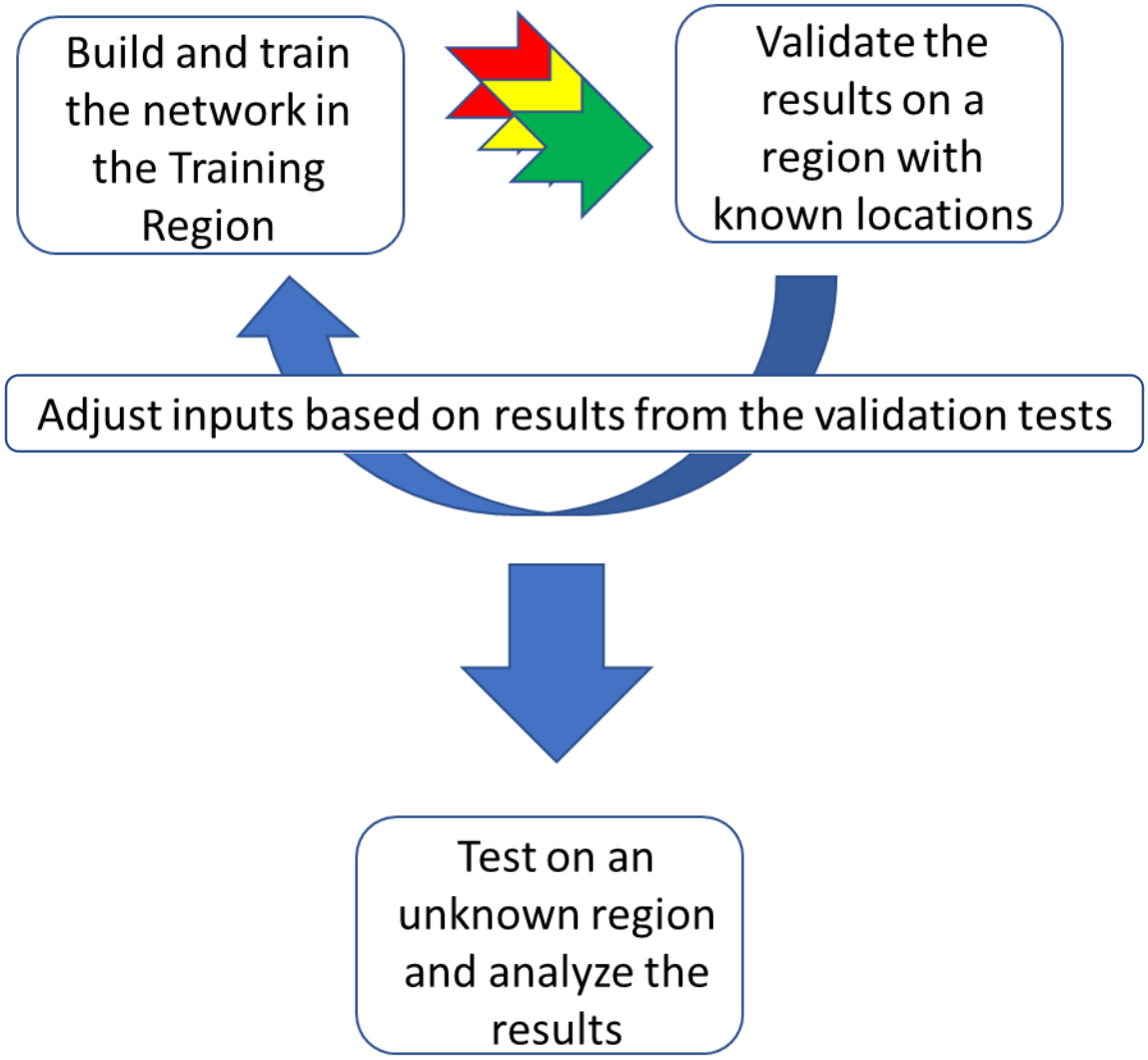


Figure 6: A neural network is trained on a region with a large sample number of a known variable before being validated in a separate area to ensure the network is working. From the validation, inputs are adjusted for better results and finally the network is applied to an entirely unknown area where results are analyzed as the final output of the network.

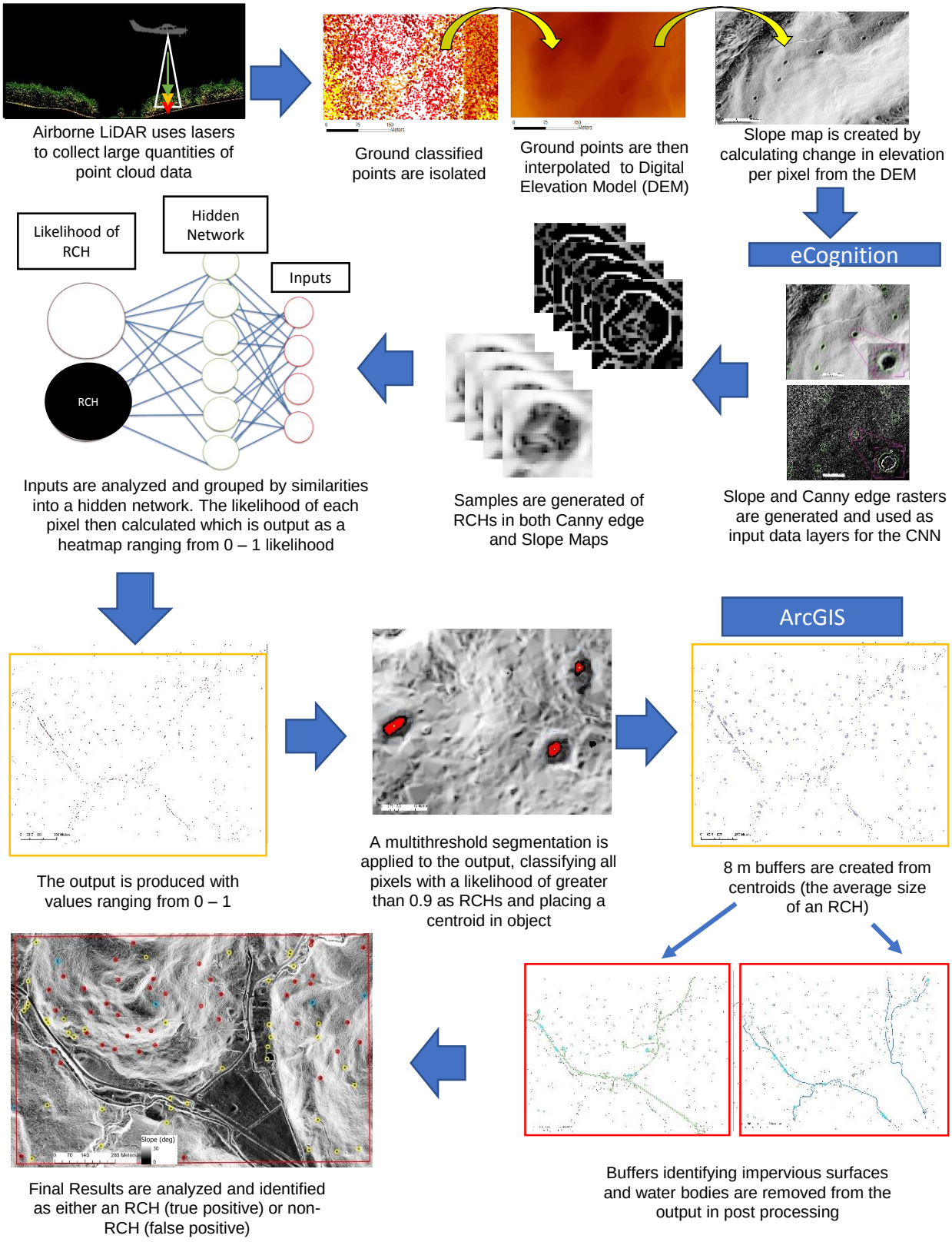


Figure 7: A schematic detailing the approach used for RCH identification

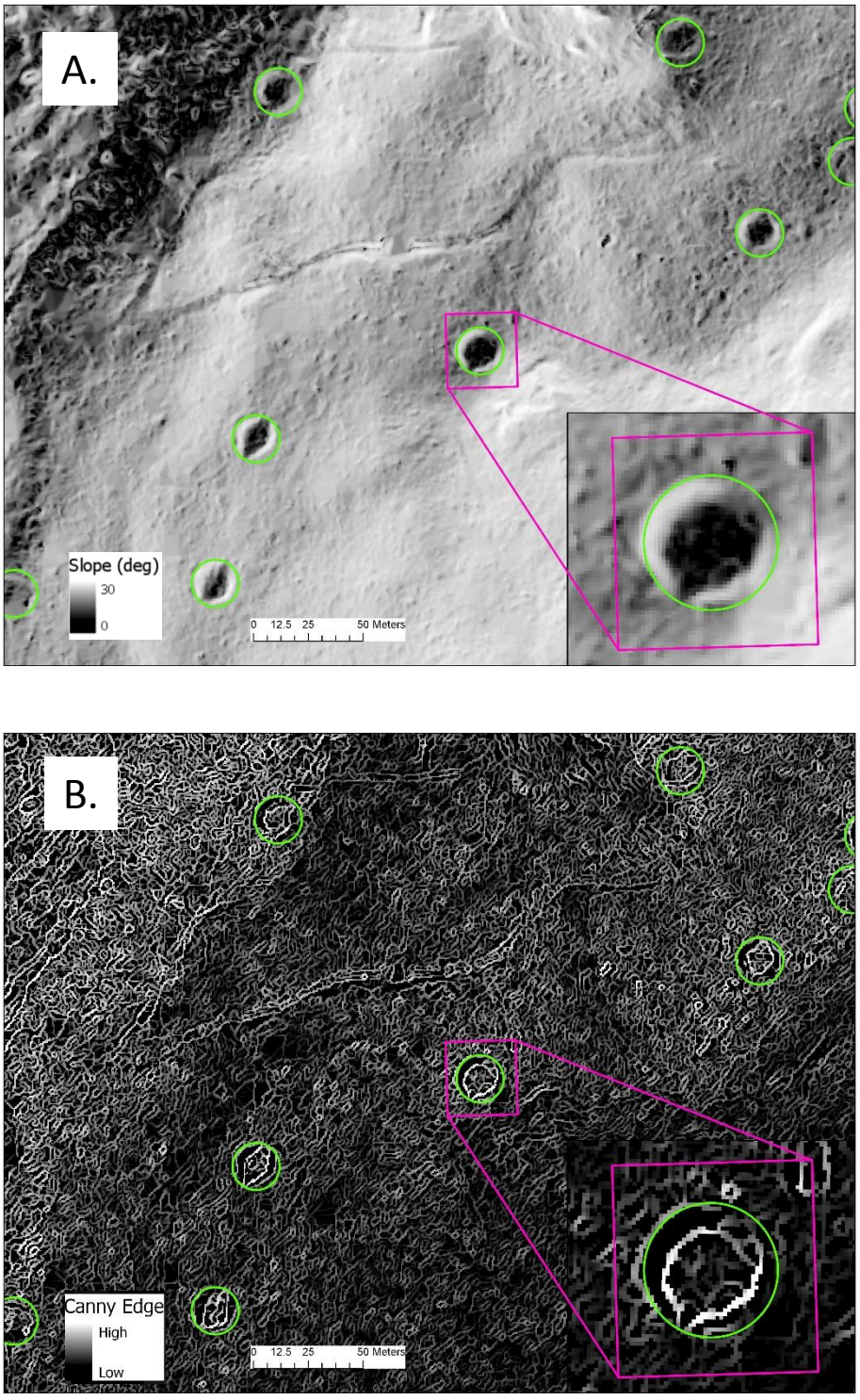
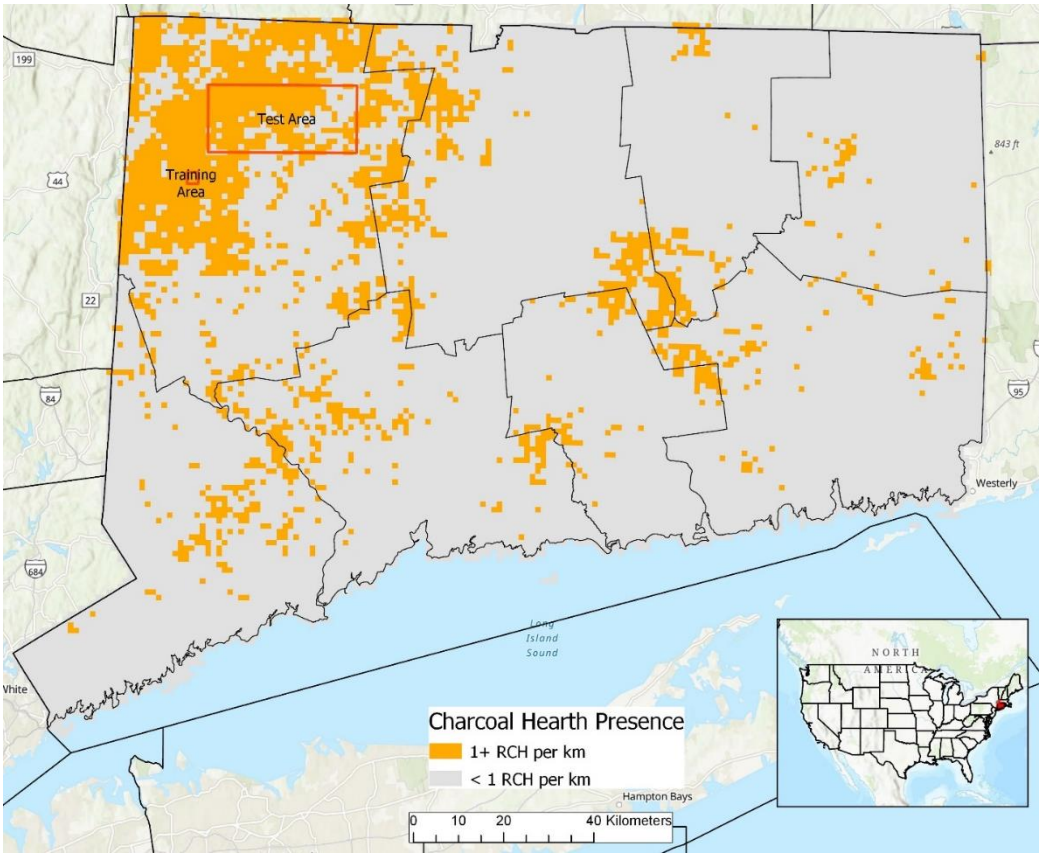


Figure 8: Slope(A) and Canny edge (B) rasters are the inputs for the network with RCHs identified in green, which were used as the inputs into the neural network. The slope highlights the change in elevation per pixel and while the edge raster highlight areas of abrupt change.



Training Region

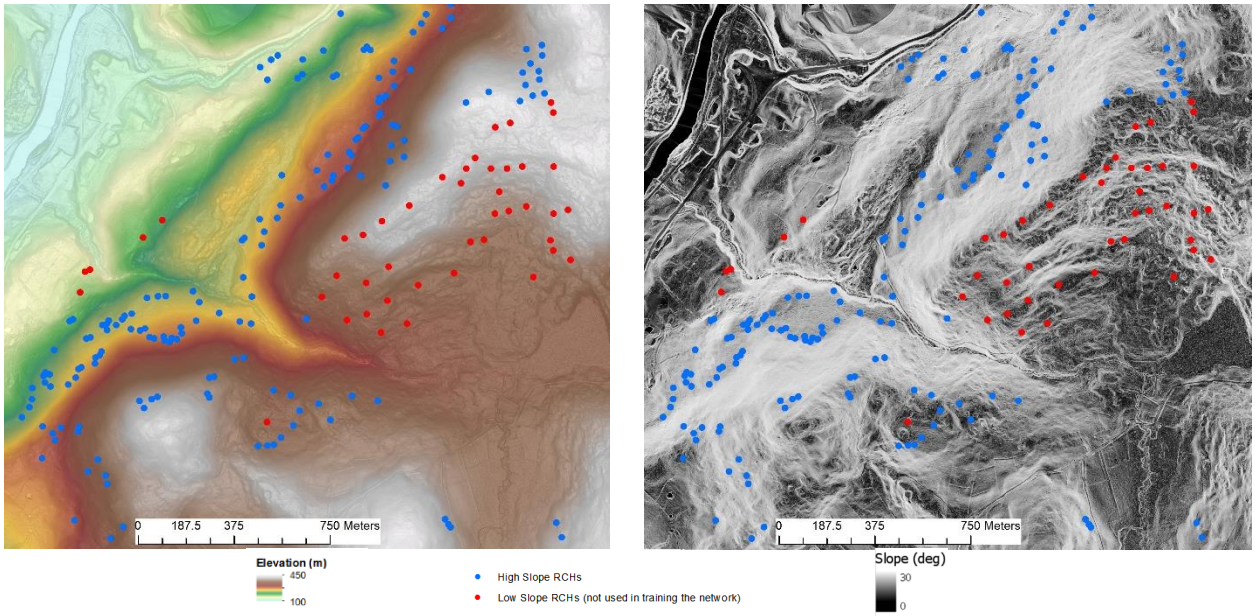


Figure 9: Test region and training region overlain on the presence/absence map of Connecticut. Below the training region is pictured demonstrating high and low slope RCHs. Elevation (A.) and slope (B.) of the training region are shown with sloped RCHs (BLUE) and near zero slope RCHs (RED) shown within the region. Only the higher sloped RCHs in BLUE were used in the development of the network.

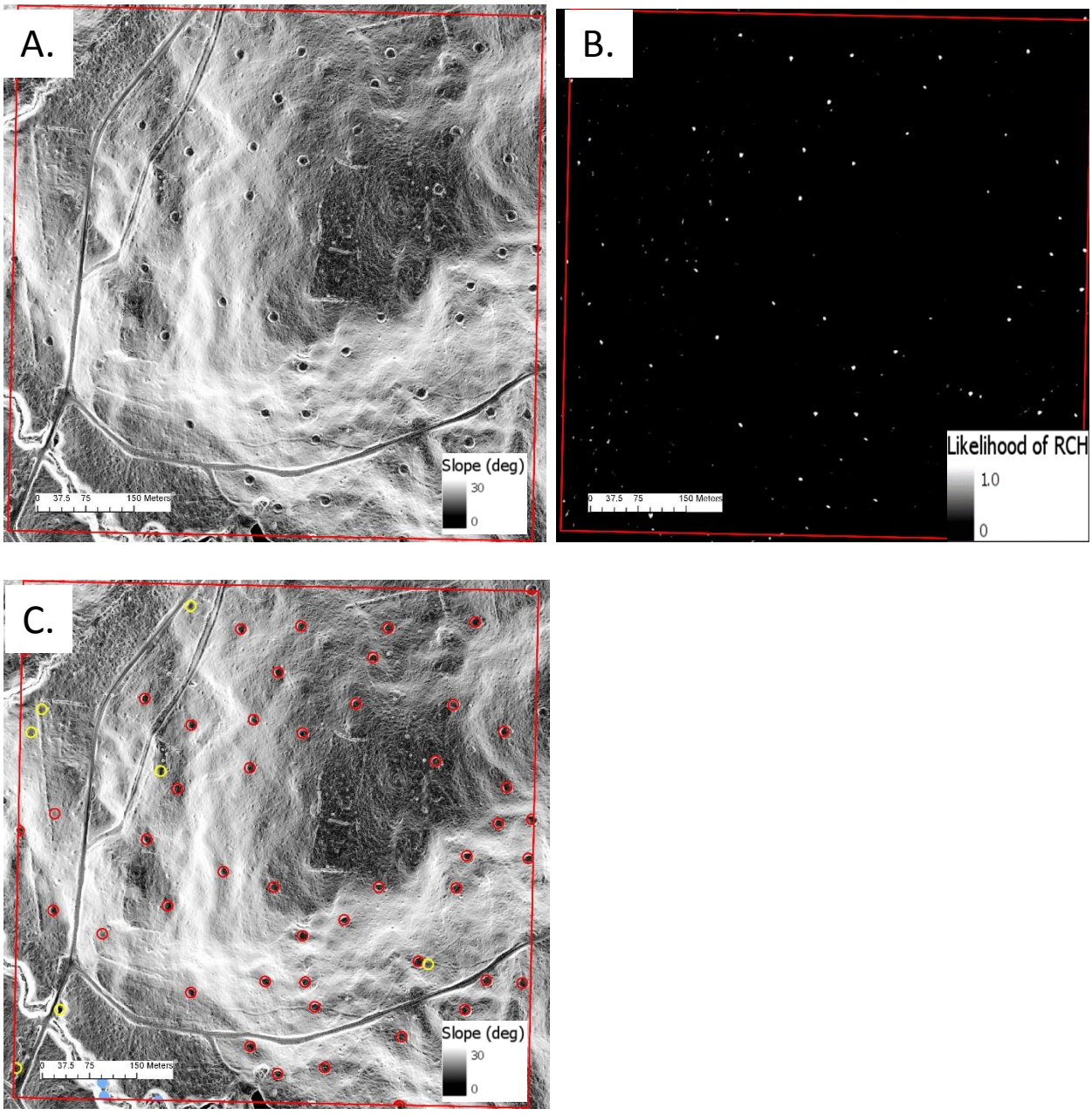


Figure 10: The validation area shown as the initial slope map input (A.), the smoothed output heatmap (B.) and the buffered results after image objects are created from groupings of 0.9 likelihood pixels (C.). There is a 90% correctness value for this region after the results are smoothed post processing removes the buffers which intersect the hydrology layer.

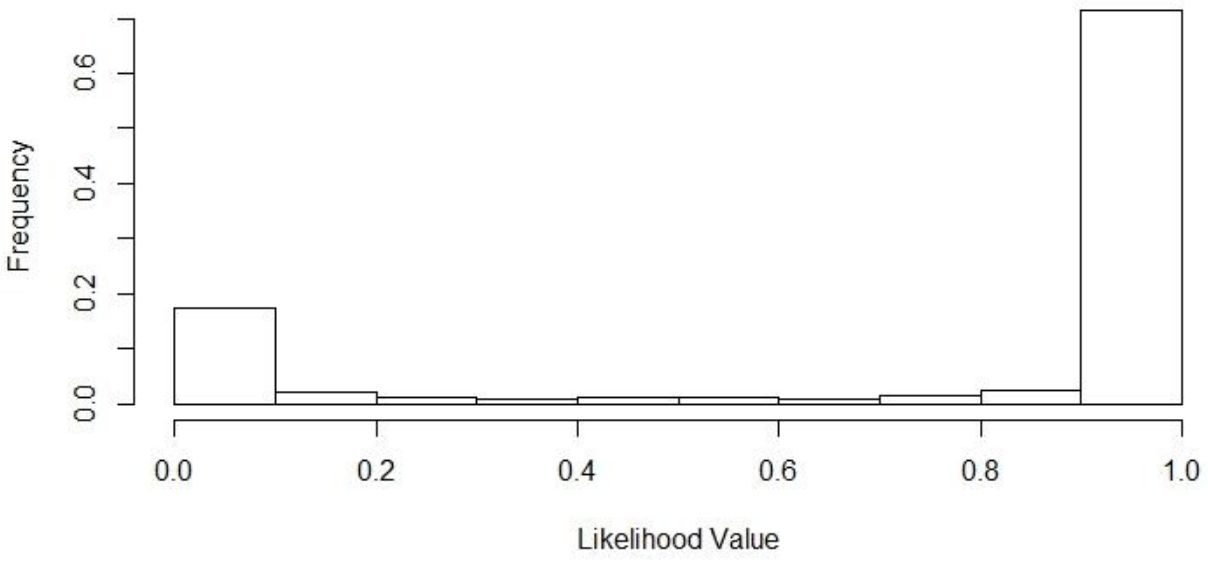


Figure 11: Relative frequency histogram of the distribution of likelihood values. 0 – 0.1 and .9 – 1.0 are significantly higher results implying that any likelihood generated is likely to be a true positive RCH.

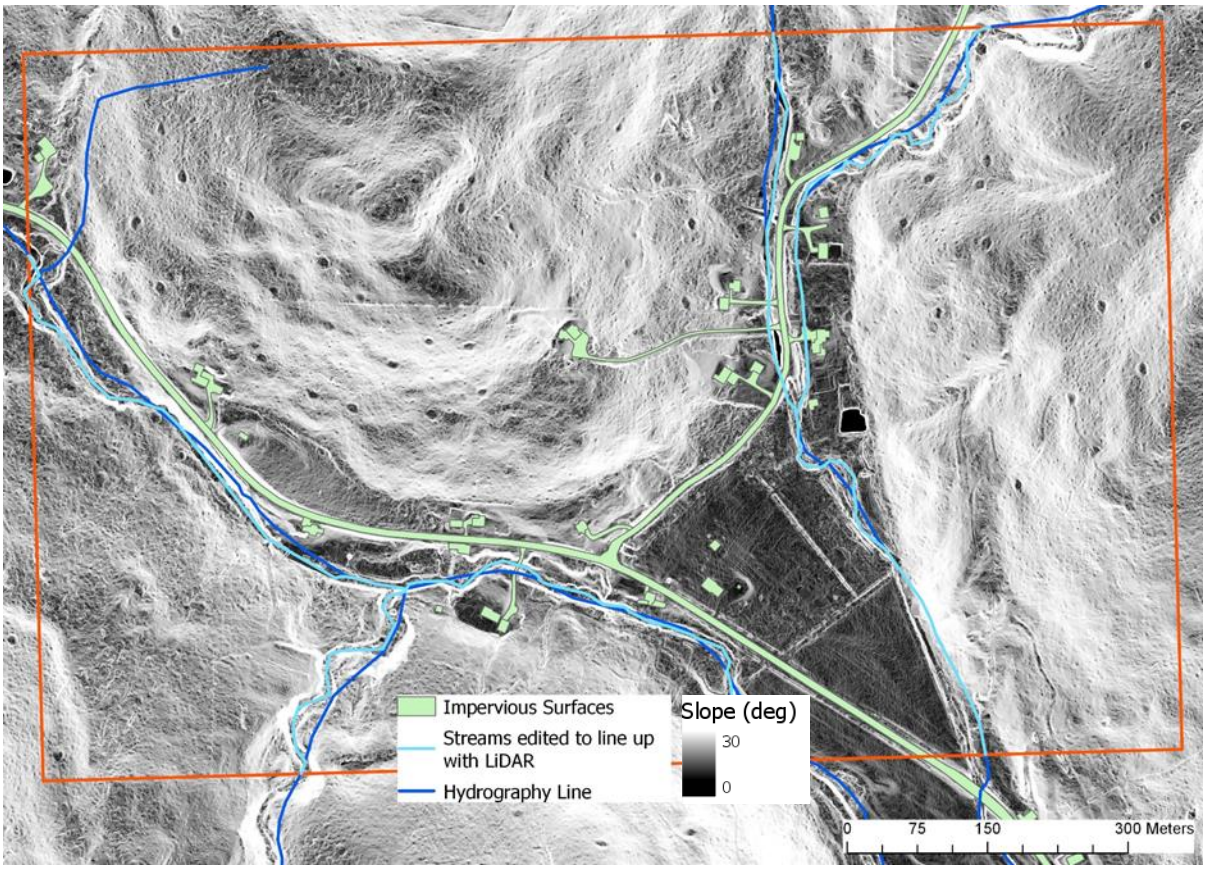


Figure 12: This figure demonstrates the use of the impervious surfaces layer, as well as showcasing the advantage in accuracy of LiDAR based hydrology over using the available datasets when applied to Test area 1.

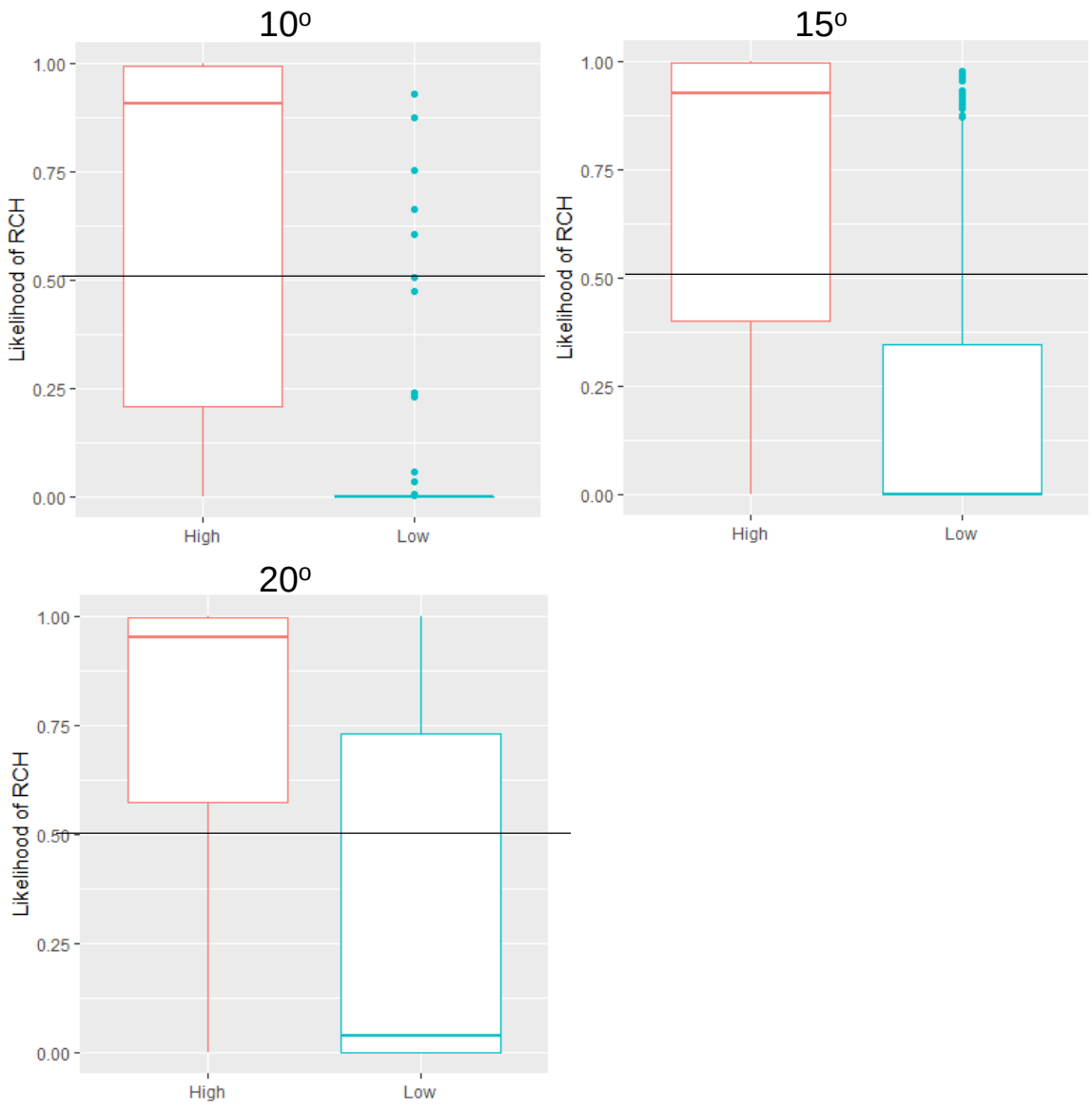


Figure 13: Figure 13 demonstrates the result of diving the high and low slopes at different slope values. The results point to high samples as 20 degrees or higher based on the clear separation between high and low slope samples above a 0.5 likelihood (Black line). This maximizes differentiation while minimizing the false positives

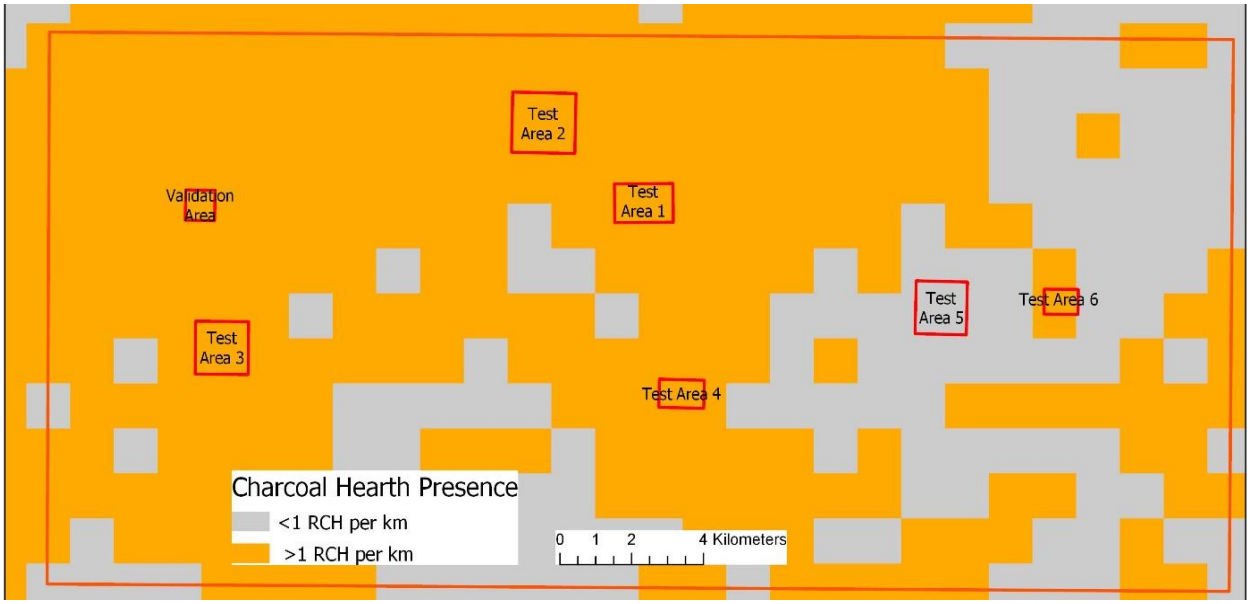


Figure 14: Test regions overlain on the RCH presence/absence grid of Connecticut

Table 1: Table 1 outlines the size, landscape type and RCH count of the validation and test areas

Test Region	Area (km^2)	Landscape Type	RCH Count
Validation Area	0.37	>15° slope with streams and paths running through the area	54
Full Test Area	277.5	Combination of all landscapes	4175
Test Area 1	0.94	>15° slopes with developed regions and stream/river bed running through the area	49
Test Area 2	1.59	Heavily developed region interspersed with >15° slopes	27
Test Area 3	1.17	Bimodal landscape with >15° and valleys	92
Test Area 4	0.53	<15° slopes with wetlands	12
Test Area 5	1.13	Fully developed with no RCHs	0
Test Area 6	0.35	>15° slopes but very rough terrain	11

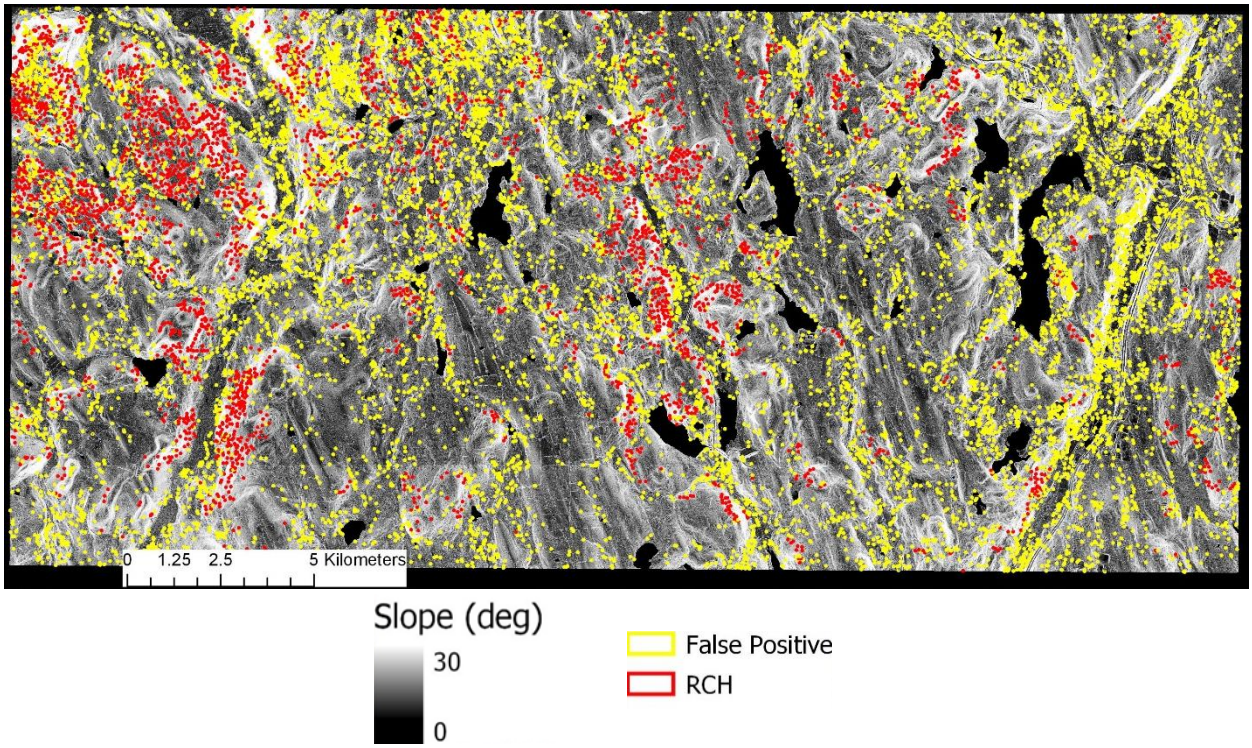


Figure 15: The above map shows the results across the entire region demonstrating why a blind application of the network to a large region is a poor idea.

Table 2: Table 2 shows the results for the entire test region after post processing.

Slope (deg)	True Positives	False Positives	Total	Recall	Precision	F1 Score
Total	3409	12435	4175	0.82	0.22	0.34
<15	65	565	337	0.19	0.10	0.13
>15	3344	11870	3841	0.87	0.22	0.35
>20	3059	10597	3206	0.95	0.22	0.36

Table 3: Results of the Validation Area and Specific Test Areas Within the Dataset

Test Area	True Positives	False Positives	Total Hearths	Recall	Precision	F1 Score
Validation Area	51	7	54	0.94	0.88	0.91
1	44	40	49	0.90	0.52	0.66
2	22	74	27	0.81	0.23	0.36
3	74	7	92	0.80	0.91	0.86
4	1	8	12	0.08	0.11	0.10
5	0	26	0	N/A	0	N/A
6	10	55	11	0.91	0.15	0.26

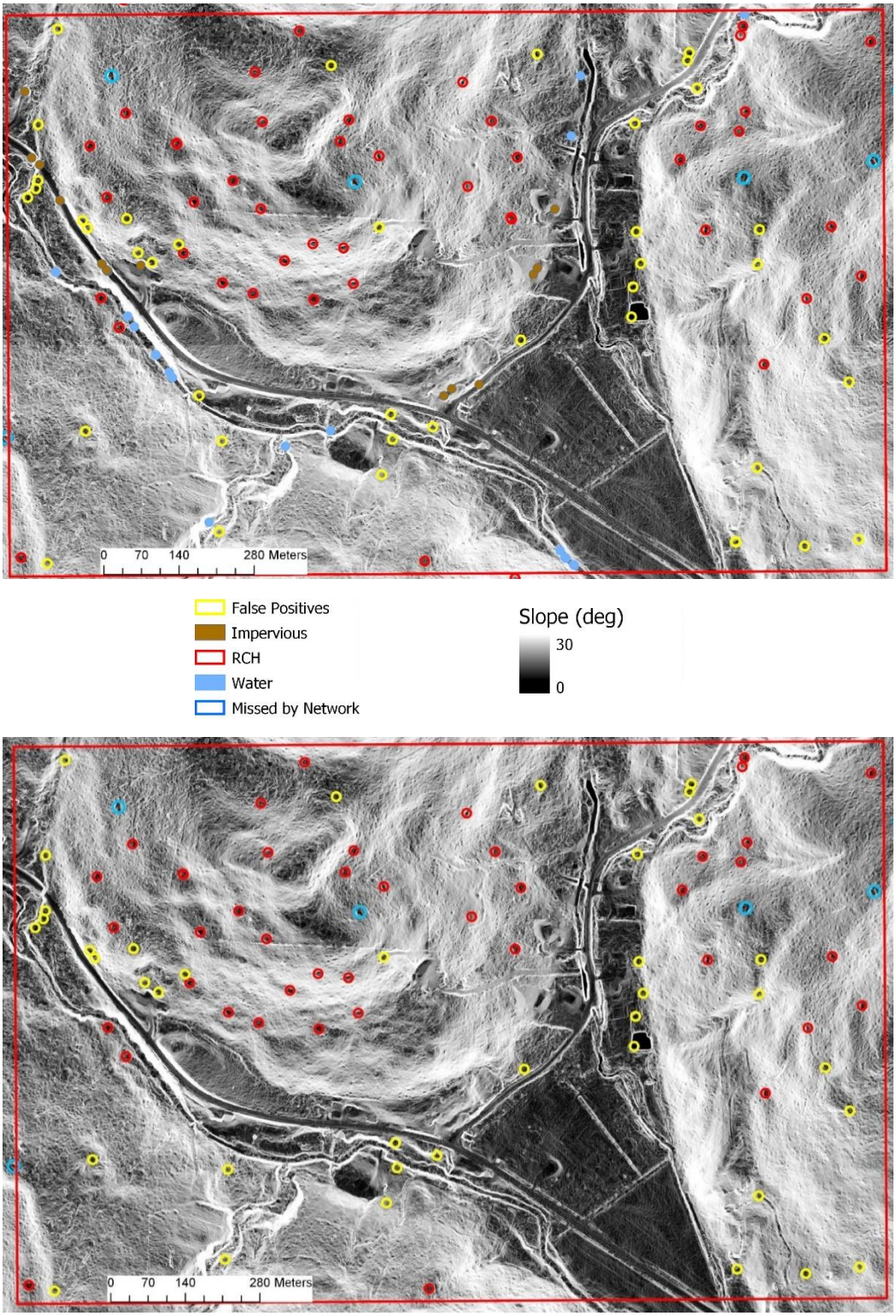


Figure 16: Test Area 1 during processing in ArcGIS and then the final result showing where impervious surfaces and hydrology layers removed false positives.

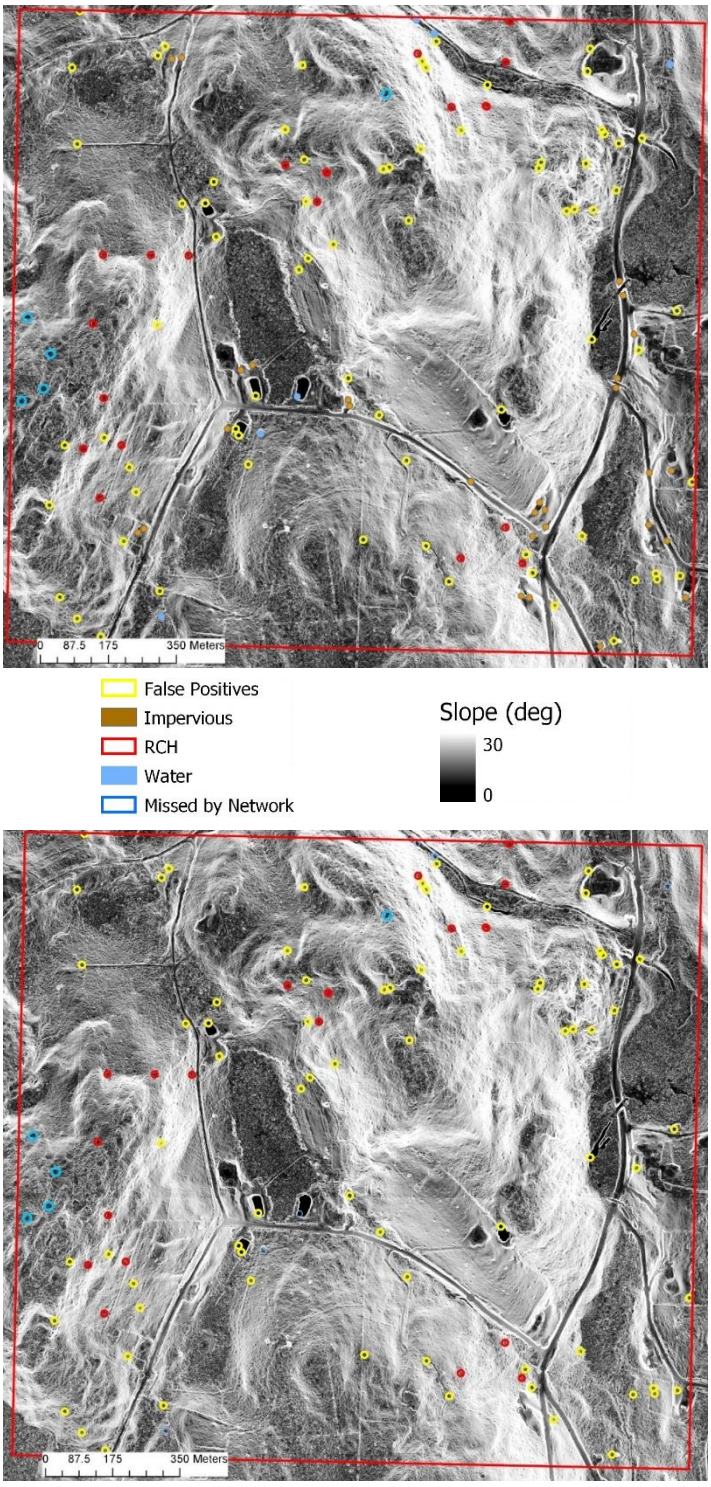


Figure 17: Test area 2 This figure shows how the results are processed using the impervious surfaces layer. It also highlights where false positives occur where buildings were built passed 2012 or the edges of water which are not identified by the waterbodies layer

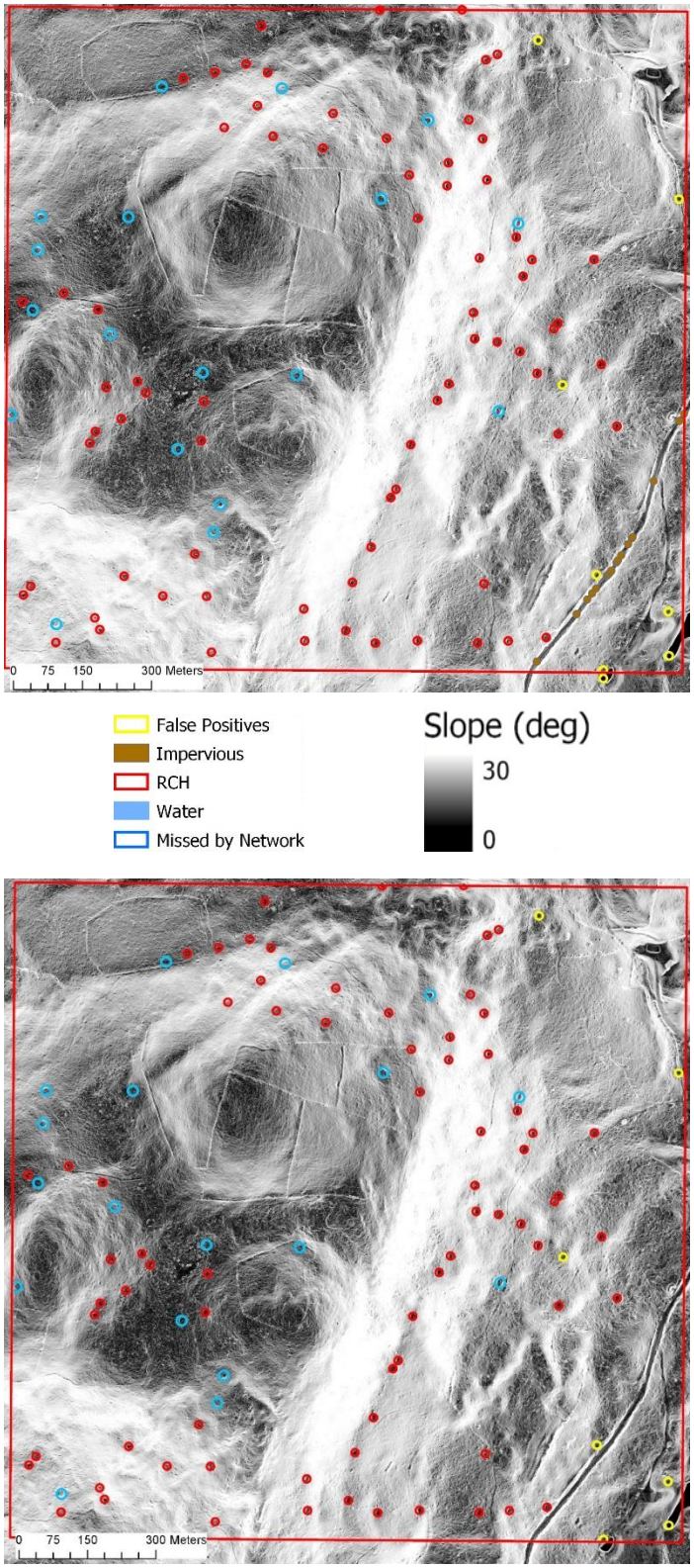


Figure 18: Test area 3 demonstrates an area where there is a mixed topography. On the sloped regions, the identification rate is close to 90%, while on near 0^{o} slopes the identification rate is closer to 50%

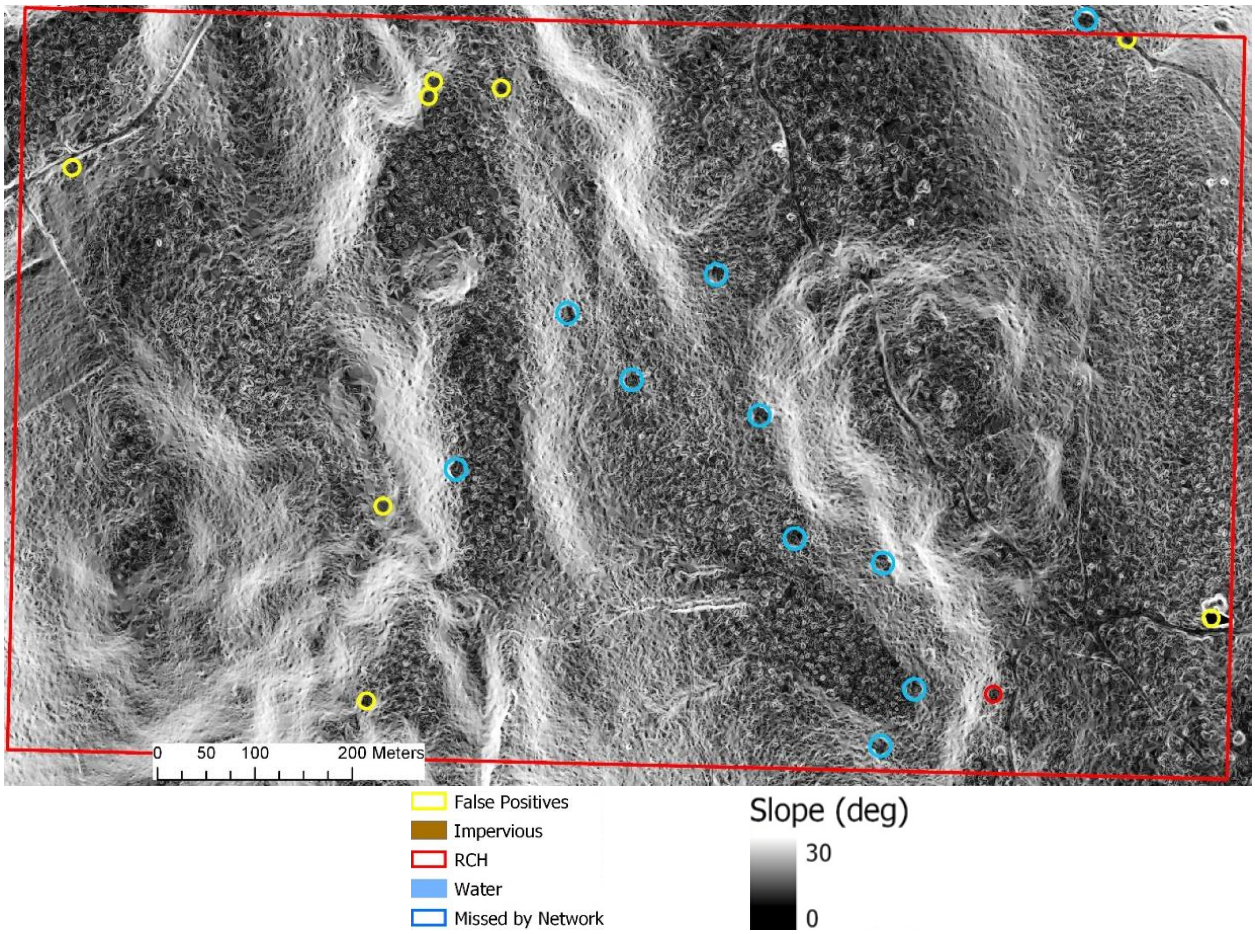


Figure 19: Test area 4 Low slope samples are hard for the algorithm to identify as the network was trained primarily high slope RCHs. False positives are both located in water which was too small for the waterbody layer to identify and remove

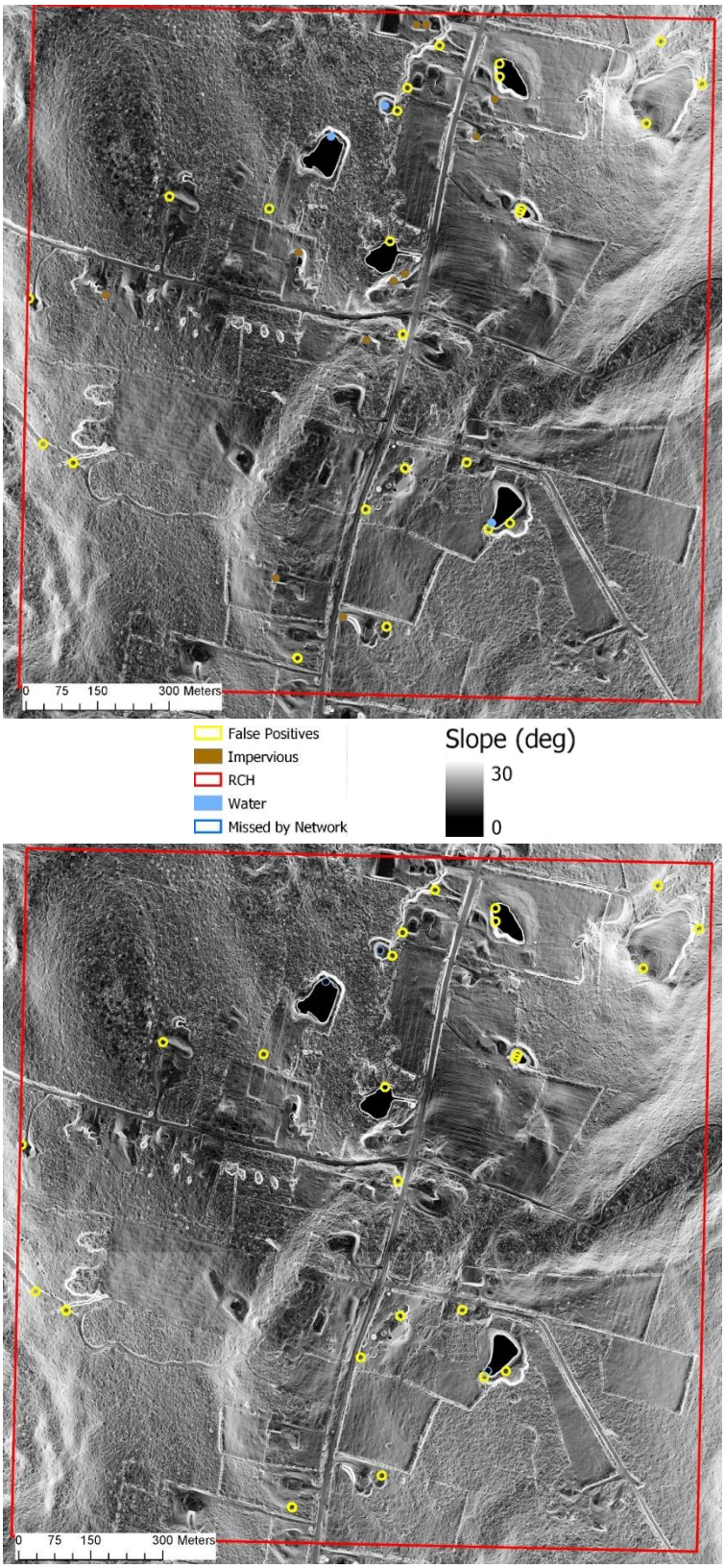


Figure 20: Test area 5 is fully developed with no true positives within the area. This highlights how many errors arise from fully developed regions

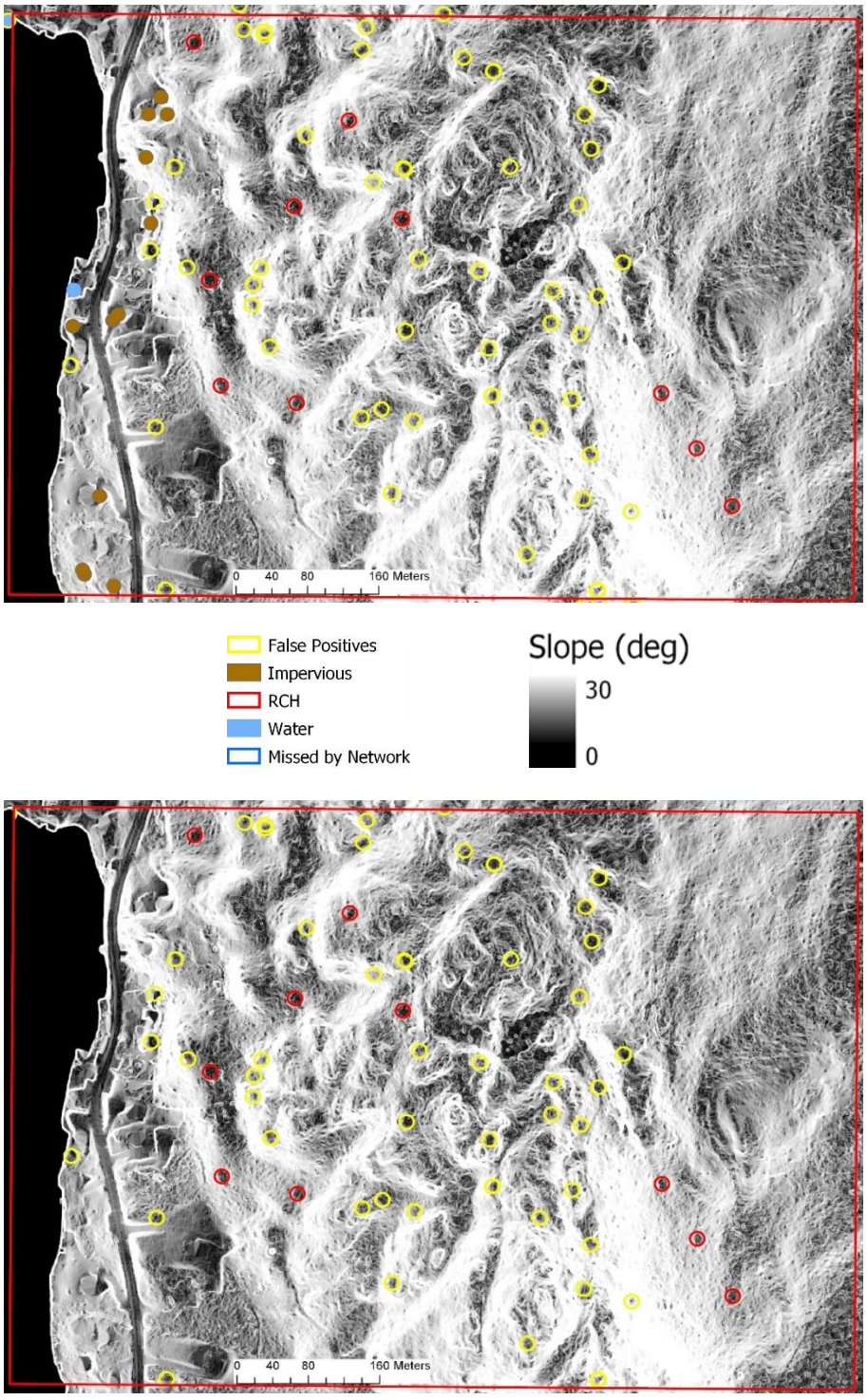


Figure 21: Test area 6 contains rough terrain, which causes issues with large numbers of false positives, despite a near 100% recall value, there are nearly 6 times more false positives than true positives leading to a precision value of only 15%.

References

1. Barger, Lucas C. (2013). *Life on a Rocky Farm: Rural Life Near NYC in the Late 19th Century*
2. Bennett, R., Welham, K., Hill, R.A., Ford, A., (2012). *A Comparison of Visualization Techniques for Models Created from Airborne Laser Scanned Data.* Archaeol. Prospect. 19, 41–48
3. Canny, John. (1987). *A Computational Approach to Edge Detection,* Readings in Computer Vision, Morgan Kaufmann, Pages 184-203, https://doi.org/10.1016/B978-0-08-051581-6.50024-6.
4. Capitol Region Council of Governments (CRCoG). (2016). *Connecticut Statewide LiDAR* 2016. http://www.cteco.uconn.edu/metadata/dep/document/lidarDEM_2016_fgdc_plus.htm
5. Challis, K., Forlin, P., Kincey, M., (2011). *A Generic Toolkit for the Visualization of Archaeological Features on Airborne LiDAR Elevation Data.* Archaeol. Prospect. 18, 279–289.
6. Cowley, D.C., (2012). *In with the new, out with the old? Auto-extraction for remote sensing archaeology,* in: Bostater, C.R., Mertikas, S.P., Neyt, X., Nichol, C., Cowley, D., Bruyant, J.-P. (Eds.), Proceedings of SPIE, Vol. 8532. pp. 853206-1-853206–9.
7. Connecticut Environmental Conditions Online (CTECO) – accessed http://www.cteco.uconn.edu/
8. Definiens. (2007). ECognition Reference Book. Trimble, Munich
9. Environmental Systems Research Institute (ESRI), (2019). ArcGIS Desktop Help 10.6 Spatial Analyst. http://desktop.arcgis.com/en/arcmap/10.6
10. Flannery, K.F, Egan-Anderson, E, and Ouimet, W.B (March, 2019). *Examining the effects of 18th-19th century charcoal production in soil-water repellency.* Poster session presented at the meeting of the Northeast Geologic Society of America, Portland, Maine.
11. Foster, David R.; Donahue, Brian; Kittredge, David; Motzkin, Glenn; Hall, Brian; Turner, Billie; and Chilton, Elizabeth S. (2008). *New England's Forest Landscape: Ecological Legacies and Conservation Patterns Shaped by Agrarian History*. Agrarian Landscapes in Transition: Comparisons of Long-Term Ecological and Cultural Change. 344. https://scholarworks.umass.edu/anthro_faculty_pubs/344
12. Foster, David and Aber, John. (2003*). Forests in Time: The Environmental Consequences of 1000 Years of Change in New England*
13. Foster, D., Motzkin, G. & Slater, B. (1998) *Ecosystems.* 1:96. https://doi.org/10.1007/s100219900008
14. Gordon, R.B., (2001). *A Landscape Transformed: The Ironmaking District of Salisbury, Connecticut.* Oxford University Press, New York.
15. Gordon, R.B., Raber, M., (2000). *Industrial Heritage in Northwest Connecticut*. Connecticut Academy of Arts and Sciences, New Haven, CT.
16. Hall, Brian; Motzkin, Glenn; Foster, David; Syfert, Mindy; and Burk, John. (2003). *Three hundred years of forest and land-use change in Massachusetts, USA*. US. Journal of Biogeography, 29, 1319–1335
17. Hesse, R., (2010). *LiDAR-derived Local Relief Models - a new tool for archaeological prospection*. Archaeol. Prospect. 17, 67–72.
18. Hirsch et al (2017). *Soils on Historic Charcoal Hearths: Terminology and Chemical Properties*. Soil Science Society of America Journal. 81. 10.2136/sssaj2017.02.0067.
19. Images from Collection of the Cornwall Historical Society, Cornwall, CT. Used with permission
20. Johnson, K.M and Ouimet, W.B., (2014). *Rediscovering the lost archaeological landscape of southern New England using airborne light detection and ranging (LiDAR)*. J. Archaeol. Sci., 43 (2014), pp. 9-20, 10.1016/j.jas.2013.12.004

21. Johnson, K. M., and W. B. Ouimet. (2016). *Physical properties and spatial controls of stone walls in the northeastern USA: Implications for Anthropocene studies of 17th to early 20th century agriculture*. Anthropocene 15:22–36.
22. Johnson, K.M., W. Ouimet, and Z. Raslan. (2015). *Geospatial and LiDAR-based analysis of 18th to early 20th century timber harvesting and charcoal production in southern New England.* Presented at the Geological Society of America Northeastern Section 50th Annual Meeting, Bretton Woods, NH. 23–25 Mar. 2015
23. Johnson, K.M. and W. Ouimet. (In Review). *Reconstructing historic forest cover and land use dynamics in the northeastern USA using geospatial analysis and airborne LiDAR*
24. Journal of the United States Association of Charcoal Iron Workers 6(1). February 1885.
25. Kirby, E., (2011). *Salisbury Iron Forged Early Industry [WWW Document].* ConnecticutHistory.org. URL http://connecticuthistory.org/salisbury-iron-forged-early-industry/ (accessed 5.12.16).
26. Kirby, E., (1998). *Echoes of Iron in Connecticut's Northwest Corner*. Sharon Historical Society, Sharon, CT.
27. Kokalj, Ž., Zaksek, K., Ostir, K., (2011). *Application of sky-view factor for the visualisation of historic landscape features in lidar-derived relief models*. Antiquity 85, 263–273.
28. Lasaponara, R., Masini, N., (2009). *Full-waveform Airborne Laser Scanning for the detection of medieval archaeological microtopographic relief.* J. Cult. Herit. 10, e78–e82.
29. Massachusetts Bureau of Geographic Information (MassGIS) - accessed https://www.mass.gov/get-massgis-data
30. McCoy, M.D., Asner, G.P., Graves, M.W., (2011). *Airborne lidar survey of irrigated agricultural landscapes: an application of the slope contrast method*. J. Archaeol. Sci. 38, 2141–2154.
31. New York State Elevation Data (NYGIS) – accessed at http://gis.ny.gov/elevation/
32. Quantum Spatial Inc. (2017). Statewide Impervious 2012. http://www.cteco.uconn.edu/projects/ms4/impervious2012.htm
33. Raab, Alexandra & Bonhage, Alexander & Schneider, Anna & Raab, T & Rösler, H & Heussner, Karl-Uwe & Hirsch, Florian. (2017). *Spatial distribution of relict charcoal hearths in the former royal forest district* Tauer (SE Brandenburg, Germany). Quaternary International. 10.1016/j.quaint.2017.07.022.
34. Rolando, V.R., (1992). *200 Years of Soot and Sweat: The History and Archaeology of Vermont's Iron, Charcoal, and Lime Industries*. Vermont Archaeological Society, Burlington.
35. Schneider, Anna & Takla, Melanie & Nicolay, A & Raab, Alexandra & Raab, T. (2015*). A Template-matching Approach Combining Morphometric Variables for Automated Mapping of Charcoal Kiln Sites. Archaeological Prospection*. 22. 10.1002/arp.1497.
36. State of Connecticut, Department of Environmental Protection (CT DEEP). (2005). Connecticut Waterbody Line. http://ct.gov/deep
37. State of Connecticut, Department of Environmental Protection (CT DEEP). (2008). Streets. http://ct.gov/deep
38. Stansfield, W, and Ouimet, W.B. (March, 2019). *History, Mapping, and Physical Properties of 18th – 19th Century Relict Charcoal Hearths in Eastern Connecticut.* Poster session presented at the meeting of the Northeast Geologic Society of America, Portland, Maine.
39. Straka, T. (2014). *Historic charcoal production in the US and forest depletion: Development of production parameters.* Advances Historical Studies 3:104–114. doi:10.4236/ahs.2014.32010
40. Schwarz, G.F. (1907). *The sprout forests of the Housatonic Valley of Connecticut.* Forestry Quarterly 5:121–153.
41. Vermont Center for Geographic Information (VGCI) – accessed at https://vcgi.vermont.gov/data/lidar
42. Winer, H.I. (1955). *History of the Great Mountain Forest, Litchfield County, Connecticut*. Yale Univ., New Haven, CT.
43. Witharana, C., Ouimet, W.B. and Johnson, K.M. (2018). *Using LiDAR and GEOBIA for automated extraction of 18th-late 19th century relict charcoal hearths in southern New England*